

# An Overview of Scalar Quantization Based Data Hiding Methods

Husrev T. Sencar<sup>a,\*</sup>, Mahalingam Ramkumar<sup>b</sup>, Ali N. Akansu<sup>a</sup>

<sup>a</sup>*New Jersey Institute of Technology, Department of Electrical and Computer Engineering, Newark, NJ 07102, USA*

<sup>b</sup>*Mississippi State University, Department of Computer Science & Engineering, Mississippi State, MS 39762, USA*

---

## Abstract

In Ref. [1], Costa presented a communications framework that provided useful insights into the study of data hiding. We present an alternate and equivalent framework with a more direct data hiding perspective. The difference between the two frameworks is in how channel dependent nature is reflected in optimal encoding and decoding operations. The connection between the suggested encoding/decoding scheme and practical embedding/detection techniques is examined. We analyze quantization based embedding/detection techniques in terms of the proposed framework based on three key aspects. The first aspect is the type of postprocessing utilized at the embedder (*i.e.* distortion compensation [2,3], thresholding [4], Gaussian mapping [5]). The second key aspect is the form of demodulation used at the extractor. The third is the criteria used to optimize the embedding/detection parameters. The embedding/detection techniques are compared in terms of probability of error, correlation, and mutual information (hiding rate) performance merits.

*Key words:* Data hiding, embedder/detector, quantization, postprocessing, thresholding, distortion compensation, Gaussian mapping.

---

\* Corresponding author.

*Email addresses:* [taha.sencar@njit.edu](mailto:taha.sencar@njit.edu) (Husrev T. Sencar), [ramkumar@cse.msstate.edu](mailto:ramkumar@cse.msstate.edu) (Mahalingam Ramkumar), [ali@oak.njit.edu](mailto:ali@oak.njit.edu) (Ali N. Akansu).

## 1 Introduction

The study of data hiding (watermarking) tries to establish the achievable limits and the design of methods for conveying a message signal, embedded within a host (cover) signal, in an imperceptible and reliable way. One conservative assumption in data hiding is that the embedder has no access to the host signal (oblivious data hiding). Though, not all data hiding applications are necessarily oblivious, our focus is the oblivious one.

The theory of data hiding has been developed mainly through employing analytical tools of communication theory. This is achieved by reinterpreting and adapting basic concepts such as channel, side information, and power constraints within the context of data hiding. In data hiding, channel is the medium between the hider and extractor, and it includes all forms of disturbances that affect the *stego* signal, which is an *intelligent* combination of the host signal and the message to be conveyed. Side information available at the encoder in a communication channel model, is associated with the host signal at the embedder in the equivalent data hiding model. Similarly, encoder/decoder pair is functionally equivalent to embedder/detector pair. Power constraints in a channel communication scenario are analogous to the perceptual distortion limits that are determined based on the features of the host signal. The bandwidth is somewhat dual to embedding signal size, and signal to noise ratio (SNR) measure corresponds to embedding distortion to attack distortion ratio (WNR) measure. Table 1 shows the relationship between the communications and data hiding frameworks. Aside from the analogies between the two frameworks, the analytical formulation of data hiding problem further requires consideration of the interactions between information hider/extractor and attacker.

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

The design principle that governs the operation of the embedder/detector pair is the most important characteristic of a data hiding method. Among a variety of research approaches the ones that draw a lot of attention are inspired from *communication with side information* [2,6–8]. Costa in [1] introduced the notion that, in a communication channel, a side information *available to encoder but not to decoder* does not necessarily causes a reduction in the communication rate. His results, when evaluated within data hiding context, encouraged researchers in designing practical oblivious data hiding schemes

Table 1  
Relationship Between Communications and Data Hiding Frameworks

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

that can achieve the hiding capacity.

To achieve the hiding rates that are closer to the upper capacity bound, several implementations are proposed [2,9,4,3,10]. These techniques are characterized by the use of enhanced quantization procedures in order to design embedding/detection techniques that approximate the performance of optimal encoding/decoding. In this class of methods, the optimal implementation requires higher dimensional quantization for embedding. In [11], Zamir *et al.* show that nested lattices can be used to construct optimum codes. However, a satisfactory performance is also achievable through scalar quantization or uni-dimensional lattices. On the other hand, the extraction of the hidden message is achieved, most generally, by employing minimum distance decoding due to the use of lattice structures in embedding.

Chen *et al.* in [9] provide a formal treatment of data hiding methods that use quantizers to embed signals, that is called quantization index modulation (QIM). In this class of methods, quantization is used to force the host signal coefficients to take desired values depending on the information signal to be embedded. Similarly, Chou *et al.* in Refs. [7,10], based on a duality with distributed source coding problem, implemented the exhaustive codeword generation of Costa's scheme by using a robust optimization method through the utilization of trellis coded quantizers. Although this approach approximates the optimal encoding/decoding scheme, even the simplest implementations involve considerable complexity. Such complexity concerns draw attention to practical approaches that handle codeword generation by scalar quantization, and therefore, have low implementation complexity

In this research direction, the most popular embedding/detection technique is a low complexity implementation of QIM which relies on uniform scalar quantization, that is called dither modulation (DM) [12]. In fact the earliest data hiding methods [13–16], which modified only 1 or 2 least significant bits (LSBs) of the host signal, are based on the same principle in rejecting the host signal interference, so called low bit modulation (LBM). For example, a method which modifies only 2 LSBs may be considered as a form of quantization index modulation where the step size of quantizer used is 4. Even-odd modulation is another embedding technique that operates similarly. In the

data hiding scheme proposed by Wang *et al.* [17], the significant wavelet coefficients are modified such that they *quantize* to an even or odd value depending on the bit to be embedded. In Ref. [18], Wu *et al.*, introduced a similar scheme based on JPEG quantizers by altering the DCT coefficients.

The performance of quantization based embedding/detection techniques, however, drops rapidly with the increase in the attack. This deficiency points out to a non-optimal design procedure compared to Costa's scheme which can deliver perfect host signal interference rejection at all attack levels. The need for a class of practical methods where the hider has better control over the operating characteristics is immediately recognized by various researchers.

In quantization-based data hiding methods, this effort resulted by incorporating a *processing stage (postprocessing) that follows the embedding quantization and by employing forms of redundancy coding*. In [2] and [12], Chen *et al.* introduced distortion compensated (DC) version of QIM (DC-QIM) (that can achieve the capacity under AWGN attacks), and spread transform (ST) technique for practical implementations (that embeds the message signal by spreading the embedding distortion over many host signal coefficients). Ramkumar *et al.* [4], considering scalar embedding, employed a thresholding type of processing at the embedder and also used a continuous triangular periodic function in order to extract the embedded binary watermark signal. In Ref. [3], Eggers *et al.* optimized the performance of DC-DM by a more careful optimization of embedding/detection parameters. They also combined multi-level signaling with binary coding techniques for low attack applications, and provided some performance results, [19]. Perez-Gonzalez *et al.* [5] proposed a probability density function (*pdf*) transformation type of processing for embedding. Furthermore, they provided a calculation of upper bound on the probability of error for multidimensional embedding case considering various noise distributions.

In this paper, we primarily concentrate on scalar quantization-based embedding/detection techniques. The main contributions are as follows:

- We present a framework for hiding methods that, from the data hiding point of view, provides a better connection between analytical results and practical designs. This interpretation enables comparison of methods on an equal footing. Accordingly, data hiding methods are studied and compared based on the following three key characteristics. These are:
  - (1) The type of the distortion reduction technique (postprocessing) employed in embedding;
  - (2) The form of demodulation used (detection function);
  - (3) The optimization criterion utilized in determining the embedding/detection parameters.
- Even though many authors have attempted to provide a comparison of

different methods, the comparison of the merits have been performed with vastly different criteria such as probability of error [9,5], correlation [4], and mutual information [3]. The framework we develop enables comparison on each or all of the criteria.

In the text following notation is used. Vectors are denoted by bold-faced characters. Random variables and their realizations are denoted by the capital and the corresponding lower case letters, respectively, in italic typeface. For the general case all signals are assumed to be vectors of size  $N$ . However, in cases where the vector random variables are independent, identically distributed (*iid*), the analysis is simplified by using the individual random variables in derivations.

In the following section, we approach the data hiding problem from *communication with side information* standpoint, where Costa's framework is revisited and an alternate framework is introduced. Extensions to practical data hiding methods are studied in Section 3. The key characteristics of quantization-based embedding/detection techniques are identified in Section 4, and the performance results are provided in Section 5. Our conclusions are given in Section 6.

## 2 Communication With Side Information and Data Hiding

Gelfand *et al.* in [20] considered a discrete memoryless channel with side information, in the form of varying channel states from a finite set, non-causally known to the encoder such that at any transmission time the encoder has the whole channel state information for all times. They proceeded to derive the capacity of this channel assuming an input alphabet  $\mathcal{X}$ , an output alphabet  $\mathcal{Y}$ , an auxiliary alphabet  $\mathcal{U}$ , and a finite set  $\mathcal{C}$  of side information where  $\mathcal{X}, \mathcal{Y}, \mathcal{U}, \mathcal{C} \in \mathfrak{R}^N$ . The channel capacity,  $C_0$ , is expressed in terms of random variables  $X \in \mathcal{X}$ ,  $Y \in \mathcal{Y}$ ,  $U \in \mathcal{U}$ , and  $C \in \mathcal{C}$  by a maximization over all conditional joint probability distributions  $p_C(c)p_{U,X}(u, x|c)p_Y(y|x, c)$  as

$$C_0 = \max_{p(u, x|c)} (I(U, Y) - I(U, C)) \quad (1)$$

where  $p_X(x)$  is the probability mass function of a random variable  $X$  and  $I(X, Y)$  is the mutual information between two random variables  $X$  and  $Y$ .

Costa [1] applied the results of [20] to memoryless channels with discrete time and continuous alphabets, and presented an information-theoretic analysis of a problem that also applies to oblivious data hiding. He studied a communications scenario where encoder transmits a message index to decoder in the

presence of a side information, and designed the auxiliary variable in Gelfand’s formulation as  $U = X + \alpha C$ , where  $X$  is the power constrained input,  $C$  is the channel state information available at the encoder, and  $\alpha$  is a scaling factor. Costa showed that for an additive white Gaussian noise (AWGN) channel with Gaussian input and side information, the channel capacity does not depend on the side information.

Later research gained considerable momentum first by reinterpreting these results in terms of oblivious data hiding, and later, by formulating the problem from a game theoretic perspective. The researchers in [21] and [22] assumed a Gaussian distributed host signal and squared error distortion measure, and studied the problem as a data hiding game between the hider/extractor and attacker. In Ref. [21], Moulin *et al.* introduced an information-theoretic model for data hiding considering memoryless attacks. In their model, the information hider determines the embedding strategy without knowing the attack, whereas the attacker uses the stego signal to design the attack. The extractor, on the other hand, is assumed to be in a position to learn the strategy of the attacker. It is shown that for squared error distortion measure and white Gaussian distributed host signal, Gaussian test channel is the optimal attack and the hiding capacity is the same as in the case where the host signal is known to the detector. They also showed that Costa’s results are valid for this setting of the data hiding game under the small distortions scenario, which assumes host signal power is much higher than that of the distortions introduced by the hider and attacker. Cohen *et al.* [22] presented a detailed discussion and the results of hiding capacity assuming Gaussian distributed host signal and squared error distortion measure, similar to [21], except for the removal of the assumption that extractor knows the attack. They showed that *iid* Gaussian host signal maximizes the hiding capacity among all finite fourth moment distributions for the host signal. Furthermore, they extended Costa’s results by considering non–white-noise attacks and non-Gaussian embedding distortions.

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

### 2.1 Costa’s Framework

Costa in [1], based on the results of [20], considered a power constrained AWGN channel with *iid* Gaussian input  $\mathbf{X}$  and side information  $\mathbf{C}$  (in the form of channel state) that is available *only* at the encoder in a non-causal

manner. A message index  $m$  is transmitted to the receiver by properly selecting the codeword  $\mathbf{X}$  that is distorted during transmission by the additive channel state  $\mathbf{C}$  and the channel noise  $\mathbf{Z}$ . Consequently, the channel output is defined as  $\mathbf{Y} = \mathbf{X} + \mathbf{C} + \mathbf{Z}$ . Considering the design of  $\mathbf{U} = \mathbf{X} + \alpha\mathbf{C}$ ,  $0 < \alpha < 1$ , and assuming  $\mathbf{X}$ ,  $\mathbf{C}$ ,  $\mathbf{Z}$  are *iid* length  $N$  sequences of random variables with zero covariance matrices and Gaussian marginal distributions (*i.e.*  $X \sim \mathcal{N}(0, P)$ ,  $C \sim \mathcal{N}(0, \sigma_C^2)$ ,  $Z \sim \mathcal{N}(0, \sigma_Z^2)$ ), the communication rate is computed as [1]

$$\begin{aligned} R(\alpha) &= I(U, Y) - I(U, C) \\ &= H(X + C + Z) + H(X) - H(X + C + Z, X + \alpha C) \end{aligned} \quad (2)$$

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

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

Hence, the rate in Eq. (2) is obtained by calculating the entropies for the corresponding distributions as [23]

$$R(\alpha) = \frac{1}{2} \log_2 \frac{P(P + \sigma_C^2 + \sigma_Z^2)}{P\sigma_C^2(1 - \alpha)^2 + \sigma_Z^2(P + \alpha^2\sigma_C^2)}. \quad (3)$$

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

The channel model for Costa's framework is displayed in Fig. 1-a. In order to transmit message  $m$ , encoder  $E$  generates the codeword  $\mathbf{X}$  that is additive to the channel state  $\mathbf{C}$  at the given channel noise variance. Decoder  $D$ , not knowing the random channel state  $\mathbf{C}$ , detects the message  $\hat{m}$  from the received signal  $\mathbf{Y}$ .

Costa outlined the capacity-achieving encoding/decoding scheme based on random coding techniques. The optimal codebook has  $M = \lfloor 2^{NR} \rfloor$ <sup>1</sup> codewords corresponding to  $M$  messages. Each message is transmitted in  $N$  uses

<sup>1</sup>  $\lfloor x \rfloor$  is the greatest integer smaller than or equal to  $x$

of the channel. For optimal encoding/decoding,  $2^{N(I(U,Y)-\epsilon)}$  (for an arbitrarily small  $\epsilon$ ) number of length  $N$  *iid* sequences with individual distributions  $\mathcal{N}(0, P + \alpha^* \sigma_C^2)$  are generated and then partitioned into  $2^{NR}$  bins. Each bin is associated with the index of a message and points to  $2^{N(I(U,C)+\epsilon)}$  number of sequences. This collection of sequences is made known to both encoder and decoder. In order to generate the codeword, the side information  $\mathbf{C}$  is weighted by the proper  $\alpha$  and subtracted from the sequences in the bin corresponding to the message to be conveyed. Among the resulting signals, the one that is orthogonal to  $\mathbf{C}$  ( $|(\mathbf{U}_j - \alpha^* \mathbf{C})^T \mathbf{C}| < \delta$ ,  $j = 1, \dots, 2^{N(I(U,C)+\epsilon)}$ , for a proper  $\delta$  value) and also satisfies the power constraint ( $\frac{1}{N} \|\mathbf{X}\|^2 \leq P$ ) is the optimal codeword corresponding to message index being sent.

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

$$|(\mathbf{U}_j - \alpha \mathbf{Y})^T \mathbf{Y}| = |(\mathbf{U}_j - \alpha^* \mathbf{C} - \alpha^* \mathbf{X} - \alpha^* \mathbf{Z})^T (\mathbf{X} + \mathbf{C} + \mathbf{Z})| \quad (4)$$

$$= |(1 - \alpha^*) \mathbf{X}^T \mathbf{X} - \alpha^* \mathbf{Z}^T \mathbf{Z}| \quad (5)$$

$$= (1 - \alpha^*) NE[X^2] - \alpha^* NE[Z^2] \quad (6)$$

$$= N \left( 1 - \frac{P}{P + \sigma_Z^2} \right) P - \frac{NP}{P + \sigma_Z^2} \sigma_Z^2 = 0. \quad (7)$$

The message index associated with the bin that contains the sequence  $\mathbf{U}_j$  is declared as the sent message. Such a code generation is asymptotically optimal as  $N \rightarrow \infty$  [1].

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

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

level.

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

$$\begin{aligned} R &= I(U, Y) - I(U, C) \\ &= H(X_n + C + Z) - H(X_n + C + Z | X + C) - H(X + C) + H(X + C | C) \\ &= H(X) + H(X_n + C + Z) - H(Z - X_t, X + C). \end{aligned} \quad (8)$$

The formulation given in Eq. (8) can be solved for rate  $R$  assuming random variables  $X$ ,  $X_t$ ,  $C$ , and  $Z$  are mutually independent except for the known dependence between  $X$  and  $X_t$ , and the variables are distributed according to  $\mathcal{N}(0, \sigma_X^2)$ ,  $\mathcal{N}(0, \sigma_{X_t}^2)$ ,  $\mathcal{N}(0, \sigma_C^2)$ , and  $\mathcal{N}(0, \sigma_Z^2)$ , respectively. The normalized-correlation between  $X$  and  $X_t$  is defined as

$$\rho = \frac{E[XX_t]}{\sqrt{E[X^2]E[X_t^2]}}. \quad (9)$$

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

$$f_{Z-X_t, X+C}(z - x_t, x + c) = \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_Z^2 + \sigma_{X_t}^2 & E[XX_t] \\ E[XX_t] & \sigma_X^2 + \sigma_C^2 \end{bmatrix} \right). \quad (10)$$

Consequently, the rate in Eq. (8) is derived by computing the entropies for the marginal and joint distributions as [23]

$$R(\sigma_X, \sigma_{X_t}, \rho) = \frac{1}{2} \log_2 \left( \frac{\sigma_X^2 (P + \sigma_C^2 + \sigma_Z^2)}{(\sigma_X^2 + \sigma_C^2)(\sigma_{X_t}^2 + \sigma_Z^2) - E[XX_t]^2} \right). \quad (11)$$

Using Eq. (9), Eq. (11) can be rewritten as

$$R(\sigma_X, \sigma_{X_t}, \rho) = \frac{1}{2} \log_2 \left( \frac{\sigma_X^2 (P + \sigma_C^2 + \sigma_Z^2)}{(\sigma_X^2 + \sigma_C^2)(\sigma_{X_t}^2 + \sigma_Z^2) - \rho^2 \sigma_X^2 \sigma_{X_t}^2} \right). \quad (12)$$

The achievable transmission rate for this channel can be found by maximizing the rate  $R$  over  $\sigma_X$ ,  $\sigma_{X_t}$ , and  $\rho$  under the constraint  $\frac{1}{N} \|\mathbf{X} - \mathbf{X}_t\|^2 = P$ . Since  $\rho$  is a normalized variable, it does not depend on the variances of  $X$  and  $X_t$ . Hence, setting  $\rho = 1$  ( $X_t$  is a linear function of  $X$ ) will maximize Eq. (12) in  $\rho$ . Moreover, the power constraint on the input relates  $\sigma_X$  and  $\sigma_{X_t}$  as

$$\sigma_{X_t} = \begin{cases} \sigma_X - \sqrt{P}, & \text{if } \rho = 1 \\ \rho\sigma_X - \sqrt{\sigma_X^2(\rho^2 - 1) + P}, & \text{if } \rho \neq 1. \end{cases} \quad (13)$$

As a result, maximization of rate given in Eq. (12) reduces to a maximization over  $\sigma_X$  for  $\rho = 1$  and  $\sigma_{X_t} = \sigma_X - \sqrt{P}$ . Then,

$$\max_{\sigma_X} R(\sigma_X, \sigma_{X_t} = \sigma_X - \sqrt{P}, \rho = 1) = \frac{1}{2} \log_2 \left( 1 + \frac{P}{\sigma_Z^2} \right) \Big|_{\sigma_X = \sigma_X^*} \quad (14)$$

which is maximized for

$$\sigma_X^* = \frac{P + \sigma_Z^2}{\sqrt{P}}, \quad \sigma_{X_t}^* = \frac{\sigma_Z^2}{\sqrt{P}}. \quad (15)$$

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

The optimal encoding/decoding scheme of the CAE-CID framework is similar to the one described in [1]. However, the encoding/decoding operations rely on the design of  $\mathbf{U} = \mathbf{X} + \mathbf{C}$  as  $\alpha$  is set to unity. Correspondingly, the shared  $\mathbf{U}$  sequences are *iid* with an underlying marginal distribution  $\mathcal{N}(0, P + \sigma_C^2)$ . The channel dependence, however, is reflected in the appropriate choice of processing that generates  $\mathbf{X}_t$  from  $\mathbf{X}$ . At the encoder, for the given  $\mathbf{C}$ , the jointly typical  $(\mathbf{U}, \mathbf{C})$  pair is searched in the bin corresponding to the message signal being sent. The codeword is generated from the  $\mathbf{U}_j$  sequence that satisfies the orthogonality constraint ( $|(\mathbf{U}_j - \mathbf{C})^T \mathbf{C}| < \delta, j = 1, \dots, 2^{N(I(U, C) + \epsilon)}$ ) and yields codeword  $\mathbf{X}_n$  such that the power constraint ( $\frac{1}{N} \|\mathbf{X}_n\|^2 \leq P$ ) is satisfied. It should be noted that, in order to achieve capacity,  $\mathbf{X}_t$  is a linear function of  $\mathbf{X}$ . Therefore, the codeword  $\mathbf{X}_n$  is readily obtained from the encoder output  $\mathbf{X}$  by the relation  $\mathbf{X}_n = \frac{\sqrt{P}}{\sigma_X} \mathbf{X}$ .

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

$$\begin{aligned}
|(\mathbf{U}_j - \mathbf{Y})^T \mathbf{Y}| &= |(\mathbf{U}_j - (\mathbf{X} - \mathbf{X}_t + \mathbf{C} + \mathbf{Z}))^T (\mathbf{X} - \mathbf{X}_t + \mathbf{C} + \mathbf{Z})| \\
&= |\mathbf{X}_t^T \mathbf{X} - \mathbf{X}_t^T \mathbf{X}_t - \mathbf{Z}^T \mathbf{Z}| \tag{16}
\end{aligned}$$

$$= NE[XX_t] - NE[X_t^2] - NE[Z^2] \tag{17}$$

$$= N \frac{P + \sigma_Z^2}{\sqrt{P}} \frac{\sigma_Z^2}{\sqrt{P}} - N \frac{(\sigma_Z^2)^2}{P} - N\sigma_Z^2 = 0 \tag{18}$$

where  $E[XX_t] = \sigma_X^* \sigma_{X_t}^*$ , Eq. (9) for  $\rho = 1$ , is used. In CAE-CID framework, since the design of the shared variable is fixed as  $\mathbf{U} = \mathbf{X} + \mathbf{C}$ , the optimal encoding/decoding merely relies on the statistics of encoder output  $\mathbf{X}$  and its dependence on processing distortion  $\mathbf{X}_t$ .

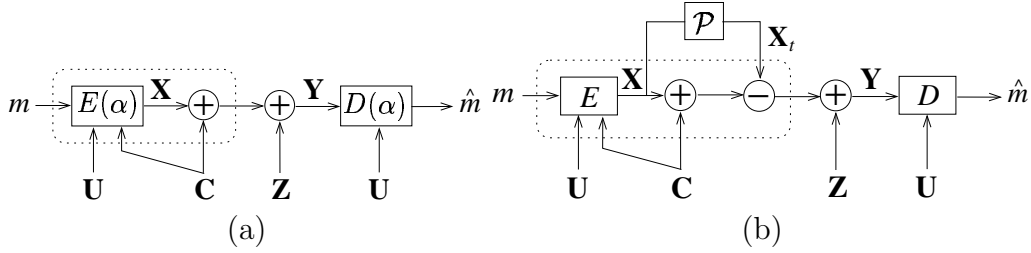


Fig. 1. The channel model for (a) Costa's framework corresponding to codebook design of  $\mathbf{U} = \mathbf{X} + \alpha\mathbf{C}$  and (b) the proposed CAE-CID framework corresponding to codebook design of  $\mathbf{U} = \mathbf{X} + \mathbf{C}$ .

When compared with Costa's framework, the CAE-CID framework has the following characteristics. In Costa's framework, the channel adaptive operation of encoder/decoder is achieved through proper selection of the scaling factor  $\alpha$ . However, in the CAE-CID framework, the channel-dependent nature of encoding is reflected in both inputs  $\mathbf{X}$  and  $\mathbf{X}_t$ . Thus, channel state interference rejection at the decoder is achieved solely by the encoder's ability to properly select  $\sigma_X$  and  $\sigma_{X_t}$  depending on the given  $\sigma_Z^2$ . Correspondingly, the CAE-CID framework provides a better theoretical basis for understanding of practical embedder/detector designs, as the postprocessing employed in practical methods can be represented by the processing distortion term  $\mathbf{X}_t$  in the formulations. In Section 4, practical embedder/detector designs will be studied from this point of view. Different types of postprocessing and their performances will be evaluated based on the choice of  $\mathbf{X}$  and  $\mathbf{X}_t$ .

### 3 Codebook Generation for Data Hiding Methods

Based on the communication frameworks given in Sections 2.1 and 2.2, encoding and decoding of a message index relies on proper selection of the codeword. Correspondingly, in the dual data hiding problem, the performance of an embedding/detection technique depends on the underlying codeword generation

scheme. Thus, main goal of a data hiding method is to design practical codebook and codeword generation schemes that can deliver perfect host signal interference rejection at all noise levels. In the context of data hiding, a codebook is a collection of mappings from the set of messages (to be conveyed) where each mapping, or codeword, is generated from the host signal by an *intelligent process* based on the imposed distortion constraints and the expected noise level.<sup>2</sup>

A typical data hiding system can be modeled as

$$\begin{aligned}
 \text{Embedding:} \quad & \mathcal{W} : m \longrightarrow \mathbf{W}, \\
 & \mathbf{S} = \mathcal{E}(\mathbf{C}, \mathbf{W}) \\
 \text{Attack:} \quad & \mathbf{Y} = \mathbf{S} + \mathbf{Z} \\
 \text{Detection:} \quad & \hat{m} = \mathcal{D}(\mathbf{Y}) \quad \text{or} \quad \hat{\mathbf{W}} = \mathcal{D}(\mathbf{Y}), \\
 & \mathcal{W}^{-1} : \hat{\mathbf{W}} \longrightarrow \hat{m}
 \end{aligned} \tag{19}$$

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

In the optimal encoding/decoding schemes of Sections 2.1 and 2.2, achieving channel capacity relies on adapting the codeword to the channel state at the given channel noise level (through the use of a very large number of  $\mathbf{U}$  sequences that are available both at the encoder and decoder). Fig. 2 depicts the encoding/decoding for message index  $m$ . In Costa's framework,  $0 < \alpha < 1$

<sup>2</sup> In order to better exploit the duality between the communication and data hiding frameworks, we define the *codeword* as the distortion signal introduced to the host signal. However, it should be noted that in some other formulations of data hiding problem, *codeword* is defined as the stego signal itself.

and processing distortion is zero, whereas in the CAE-CID framework,  $\alpha = 1$  and the processing distortion is non-zero. (Hence, the main difference between the two frameworks is in *how the channel dependent nature is reflected in encoding/decoding operations.*) Despite their optimality, such encoding/decoding schemes cannot be applied to the design of practical embedding/detection techniques the way they are due to complexity issues. However, the underlying design principles can be applied [2,4,10,3,5].

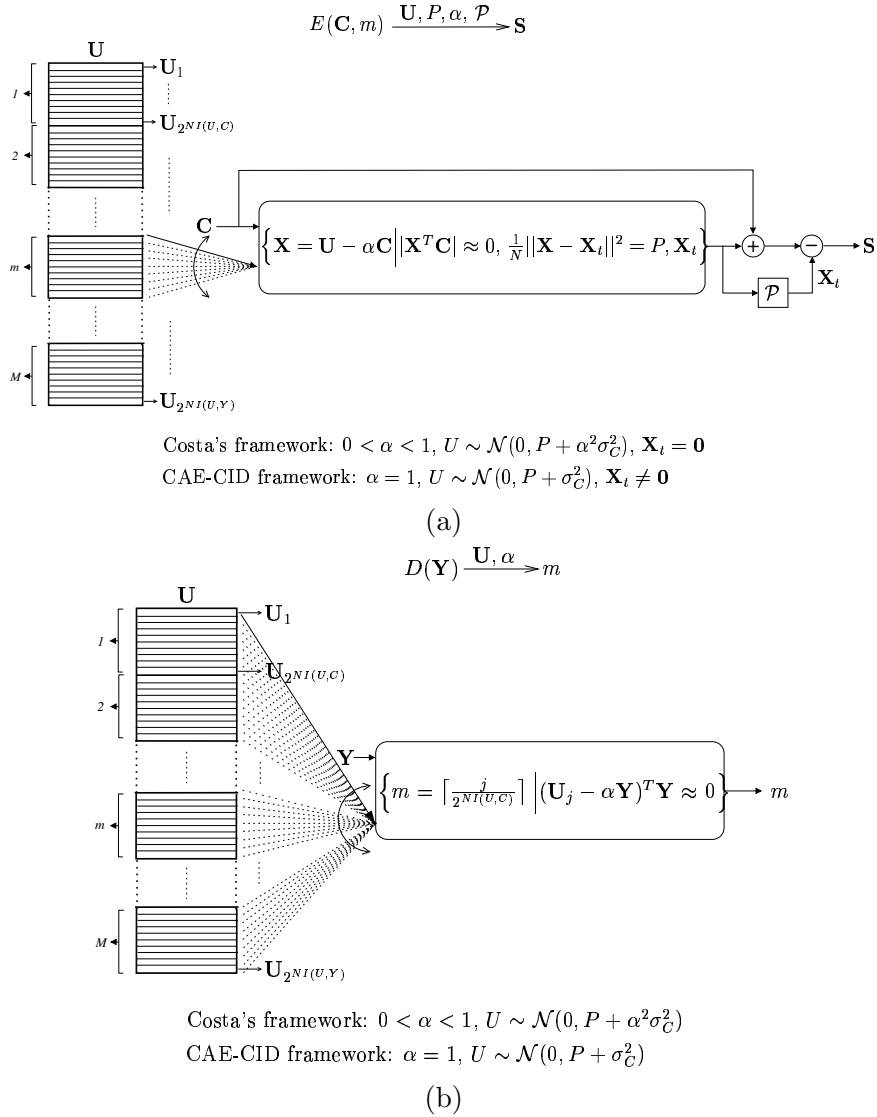


Fig. 2. Optimal (a) encoding and (b) decoding of a message index  $m$ .

An efficient algebraic structured binning scheme that generalized the approach for constructing optimum codes for data hiding is provided by [11,24,8]. This approach connects the embedder/detector design problem with the areas of linear codes and lattice codes. A nested lattice code is essentially a high-dimensional lattice partition characterized by a *fine* lattice and a *coarse* lattice.

The fine lattice is partitioned into a number of cosets corresponding to *coarse* lattice and its translates. (In other words, the union of the coarse lattice and its translates is the fine lattice.) Hence, the coding rate is determined by diluting the coset density in the *fine* lattice space. Accordingly, the encoding of the message  $m$  is performed by quantizing  $\mathbf{C}$  in the  $N$ -dimensional signal space to the nearest lattice point in the corresponding *coarse* lattice, and the decoding is by quantizing the received  $\mathbf{Y}$  to the nearest point in the *fine* lattice. Similarly, the embedding rate is designated by the number of cosets. The construction of good nested lattice codes corresponds to the use of high-dimensional vector quantization for embedding and detection. (It should be noted that QIM and DC-QIM [9] are constructions based on high-dimensional self-similar lattices where the coarse lattice is scaled and rotated version of the fine lattice.) However, from the practical point of view, high-dimensional constructions are not feasible. Therefore lattices with simpler structures need to be utilized. Such constructions include recursive quantization procedures and Cartesian products of low-dimensional lattices which coincide with the practical embedder/detector designs.

In quantization-based methods, the optimal encoding/decoding procedure is effectively simplified by generating  $\mathbf{U}$  sequences as sequences of reconstruction points where each reconstruction point is associated with a quantizer from a set of quantizers. The number of quantizers in the set corresponds to number of messages or message (watermark) letters. Each quantizer of the set is uniquely described by a set of reconstruction points that are non-overlapping with the other sets of reconstruction points. Therefore, each finite state of  $\mathbf{U}$  is a sequence with values restricted to reconstruction values of the designated quantizers. The crux of practical methods is that each codeword is directly generated from the given host signal and the watermark signal through quantization rather than maintaining a collection of shared  $\mathbf{U}$  sequences.

Practical data hiding approaches can be categorized into three main types, based on the frameworks studied in Sections 2.1 and 2.2, depending on the design of embedder/detector pair, namely type-I, type-II, and type-III [25,26].

Since the goal of Section 3 is to describe and explain the codebook generation for the three types of embedder/detector, we preferred to incorporate the two frameworks in order to *better* characterize and represent the codebook design as summarized in Table 2, and Section 3.1 is based on this premise. Table 2 incorporates the two frameworks to describe and characterize the three types of methods. It should be noted that due to the construction of the codebook  $\mathbf{U}$ , type-I schemes fit better into Costa's framework whereas type-II and type-III schemes can be better understood within the CAE-CID framework as will be discussed in the following subsections.

Based on the codebook designs, it is observed that type-I embedding does

Table 2  
Three Types of Embedding/Detection Schemes

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

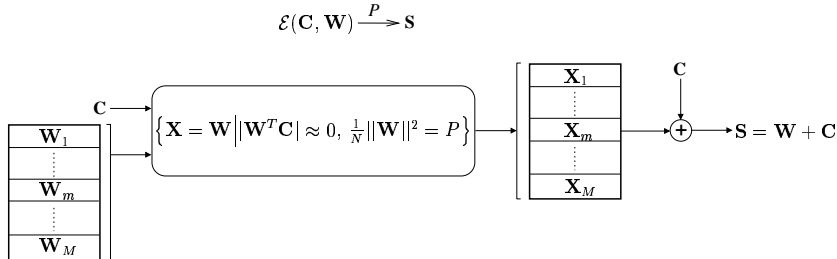


Fig. 3. Encoding of message index  $m$  in type-I methods.

not exploit any information on host signal or channel noise level. While type-II embedding exploits only host signal information. Type-III embedding, on the other hand, utilizes both forms of information. The choices of design for type-I and type-II frameworks correspond to two extreme cases in hiding rate vs. robustness curves. Namely, type-I methods are preferred for the case of “severe attacks” while type-II methods are superior for the case of “low attacks.” Whereas an optimal design imposes some sort of dependency on the channel noise instead of the fixed severe noise (type-I) or low noise (type-II) assumptions. In type-III methods, the design principle is based on maximizing the data hiding rate at the presumed noise level. Therefore, type-III is a generalization of type-I and type-II, and its optimal version achieves the data hiding capacity at all WNRs.

### 3.1 Type-I Embedding and Detection

Type-I methods refer to additive schemes where the stego signal is generated by adding the watermark signal to the host signal. The codebook generation of type-I methods corresponds to design of  $\mathbf{U} = \mathbf{X}$ ,  $\alpha = 0$ , in Costa’s framework. In the CAE-CID framework, on the other hand, corresponding design takes the form of  $\mathbf{U} = \mathbf{X} + \mathbf{C}$  when  $\rho = 1$  and  $\sigma_X^2 = \sigma_C^2 + \sigma_Z^2$ . However, in order to better emphasize the differentiation between the three types of methods, we will identify type-I methods in terms of Costa’s framework. Accordingly, the codeword is the watermark signal  $\mathbf{X} = \mathbf{W}$ , and it satisfies the power and orthogonality constraints. Fig. 3 depicts the codeword generation for type-I methods for a given set of watermark signals and the host signal.

In type-I schemes, optimal decoding of the embedded message depends on exact probabilistic characterization of the host signal at the detector. This

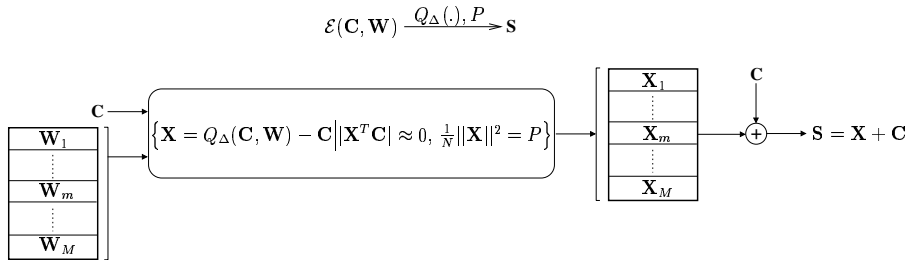


Fig. 4. Encoding of message index  $m$  in type-II methods.

type of methods suffer severely from host signal interference due to the non-optimal design that assumes the host signal  $\mathbf{C}$  as a noise and tries to cancel it. Therefore, they have preferable performance only if channel noise is very strong or the host signal is available at the extractor. In other words, data hiding rate is uncompromisingly traded off against robustness to severe attacks.

### 3.2 Type-II Embedding and Detection

The codebook generation for type-II methods is characterized by the design of  $\mathbf{U} = \mathbf{X} + \mathbf{C}$  which corresponds to choice of  $\alpha = 1$  within Costa's framework or  $\mathbf{X}_t = 0$  ( $\mathbf{X}_n = \mathbf{X}$ ) within the CAE-CID framework. The corresponding design of  $\mathbf{U}$  can be readily implemented by the use of quantization procedures. Except the codebook design, type-II methods are also characterized by their  $\mathcal{E}, \mathcal{D}$  designs, which are exact inverses,  $W = \mathcal{D}(\mathcal{E}(C, W))$ .

In type-II methods, embedding a message into a host signal refers to quantization of the host signal by a quantizer picked from an ensemble of quantizers, where each quantizer is associated with a message letter or message index. Thus, the stego signal  $\mathbf{S}$  is a quantized form of  $\mathbf{C}$ , and the corresponding quantization error, introduced to the host signal  $\mathbf{C}$ , is the codeword  $\mathbf{X}$ . The number of quantizers in the ensemble determines the information embedding rate. The embedding distortion is measured using squared error distance measure, *viz.*,  $\frac{1}{N} \|\mathbf{X}\|^2 = P$ , and it varies with the size and shape of the quantization cells. The orthogonality constraint,  $\mathbf{X}^T \mathbf{C} = 0$ , however, is relaxed by assuming that  $\mathbf{C}$  is uniformly distributed over all quantization cells and the number of quantization levels is not small such that  $\mathbf{X}$  and  $\mathbf{C}$  are approximately uncorrelated. This assumption also removes the dependence of embedding/detection operations on the host signal's statistics. In practice, this can be satisfied by the *small distortions scenario* where embedding and attack distortion powers are much less than the host signal power. The codeword generation of type-II methods is depicted in Fig. 4, where  $Q_{\Delta}(\cdot)$  is a high dimensional vector quantizer or a Cartesian product of scalar quantizers with  $\Delta$  as the distance between the reconstruction points.

A practical implementation of type-II methods is based on dithered quantizers, *viz.* dither modulation (DM). Dithered quantizers intend to decorrelate the quantization error of a quantizer from its input [27]. In subtractive dithering, an *iid* dither vector (independent of the input) is added to the input prior to quantization, and then subtracted from the quantized output. Hence, the goal (decorrelation of the quantization error with the input) is achieved. Within the context of data hiding, the dither signal is merely a mapping from the message index, the watermark signal. Therefore, the dither signal is not genuinely random and the orthogonality between the error and the input signals is not guaranteed.

In DM, each quantizer in the ensemble is generated from a base quantizer by shifting the quantization cells and reconstruction points. The stego signal is generated by quantizing the host signal with the corresponding dithered quantizer as

$$\mathbf{S} = Q_{\Delta}(\mathbf{C} + \mathbf{W}_m) - \mathbf{W}_m \quad (20)$$

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

$$\mathbf{X} = (Q_{\Delta}(\mathbf{C} + \mathbf{W}_m) - \mathbf{W}_m) - \mathbf{C}. \quad (21)$$

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

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

$$q_{\Delta}(C) = i\Delta, \quad \text{for } i\Delta - \frac{\Delta}{2} \leq C < i\Delta + \frac{\Delta}{2}. \quad (22)$$

Therefore, embedding can be viewed as  $N$  successive scalar quantization, of the coefficients of  $\mathbf{C} = (C_1, \dots, C_N)$ , dithered with the watermark signal vector  $\mathbf{W}_m = (W_{m_1}, \dots, W_{m_N})$ . Each distinct component of the watermark (dither) signal is associated with a quantizer that is generated by properly shifting the reconstruction points of  $q_{\Delta}(\cdot)$ . The amount of shifting is determined by the number of possible values a watermark sample can take (the number of quantizers). For maximum separation of the reconstruction points of embedding quantizers, the watermark sample values are equally spaced along an interval of length that is equal to quantization step size  $\Delta$ , *i.e.*,  $[-\Delta/2, \Delta/2)$ . It

should be noted that the sample values represented by the form  $W_m + i\Delta$  for  $i \in \mathcal{Z}$ , where  $\mathcal{Z}$  is the set of all integers, lead to the same dithered quantizer. (In other words, shifts differing by an integer multiple of  $\Delta$  correspond to the same quantizer.) Considering a  $d$ -ary watermark sample, the set  $\Omega$  that contains the  $d$  possible sample values is defined as

$$\Omega = \left\{ \delta + i\Delta, \delta + \frac{\Delta}{d} + i\Delta, \delta + 2\frac{\Delta}{d} + i\Delta, \dots, \delta + (d-1)\frac{\Delta}{d} + i\Delta \right\} \quad (23)$$

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

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

### 3.3 Type-III Embedding and Detection

The poor performance of type-II methods with increasing attack levels is substantially improved by enhancing the functionality of the type-II embedder with further processing capabilities (*i.e.* thresholding, distortion compensation, Gaussian mapping). In type-III methods, embedding quantization is followed by a processing stage (postprocessing) that generates the stego signal. The postprocessing is designed in a way that hiding rate is maximized at the presumed attack level. On the other hand, the invertibility condition on the  $(\mathcal{E}, D)$  pair is sacrificed as a result of the postprocessing,  $\mathcal{D}(\mathcal{E}(C, W)) \neq W$ .

Codebook design of type-III methods follows  $\mathbf{U} = \mathbf{X} + \mathbf{C}$  when  $\rho = 1$  and  $\mathbf{X}_t \neq \mathbf{0}$  within the CAE-CID framework, and  $\mathbf{U} = \mathbf{X} + \alpha\mathbf{C}$  where  $0 < \alpha < 1$  within Costa's framework. However, codeword generation for most type-III methods does not explicitly follow Costa's framework due to the processing that takes place after quantization of the host signal. Therefore, type-III methods are better evaluated within the CAE-CID framework. Fig. 5 depicts the corresponding codeword generation. In type-III methods, the quantization error (type-II codeword) undergoes the particular processing  $\mathcal{P}$  which generates

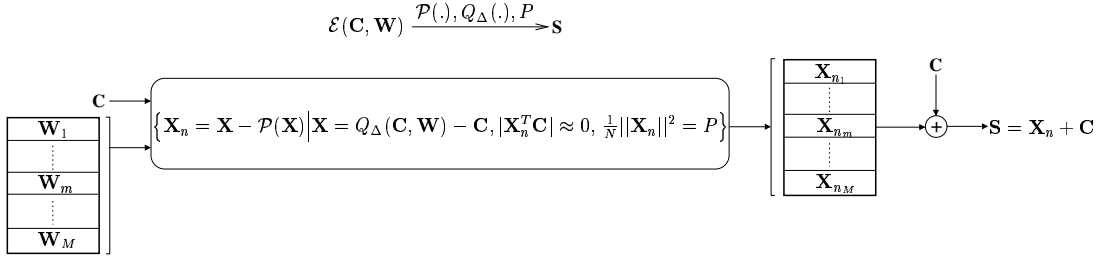


Fig. 5. Encoding of message index  $m$  in type-III methods.

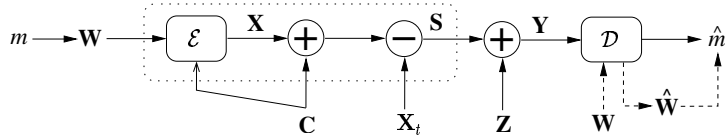


Fig. 6. Block diagram of type-III embedding and detection stages.

the codeword  $\mathbf{X}_n = \mathbf{X} - \mathcal{P}(\mathbf{X})$ .

The improvement in the performance of type-III methods, compared to type-II, at the same noise level can be explained by the fact that codebook design depends on channel noise level or by the deviation from the non-optimal design of  $\mathbf{X}_t = \mathbf{0}$  through the added processing. Alternately, in terms of Costa's framework, the improvement can be attributed to the effective value of  $\alpha$  used in codebook generation, which is less than one rather than being equal to one as the latter is optimal for the no attack case. Data hiding methods with post-processing abilities enable the embedder to increase the distance between the reconstruction points of quantizers at a fixed embedding distortion. Therefore, they have improved detection capabilities for any finite WNR level (type-II is optimal only for the case of infinite WNR). On the other hand, since the detector is blind to the additional processing at the embedder, its structure is not altered.

The channel model for type-III hiding methods is displayed in Fig. 6. In the model,  $\mathbf{X}$  is the type-II codeword (embedding distortion introduced due to the quantization),  $\mathbf{X}_t$  is the processing distortion, and the channel output is  $\mathbf{Y} = \mathbf{C} + \mathbf{X} - \mathbf{X}_t + \mathbf{Z}$ . The processing distortion  $\mathbf{X}_t$  is derived from  $\mathbf{X}$  by the postprocessing depending on the expected noise level. The type-III codeword that yields the stego signal,  $\mathbf{S} = \mathbf{C} + \mathbf{X}_n$ , is defined as  $\mathbf{X}_n = \mathbf{X} - \mathbf{X}_t$ . Correspondingly, embedder imposes the power constraint as  $\frac{1}{N} \|\mathbf{X}_n\|^2 = P$ .

In type-III methods, since the detector is not aware of the processing at the embedder, the processing distortion  $\mathbf{X}_t$  can effectively be considered to be a part of the channel noise at the detector. Therefore, type-II codeword  $\mathbf{X}$ , which would yield an errorless extraction of the watermark signal  $\mathbf{W}$ , is distorted by two sources of noise, *viz.*, the attack  $\mathbf{Z}$  and the processing distortion  $\mathbf{X}_t$ . Therefore, the effective noise at the detector that distorts the embedded

watermark signal can be represented as  $\mathbf{Z}_{eff} = \mathbf{Z} - \mathbf{X}_t$ .

## 4 Performance Evaluation

Performance and design of quantization-based data hiding methods, that rely on type-II and type-III hiding principles, vary based on three factors: the type of postprocessing incorporated with type-II embedding, the choice of demodulation function used in message extraction, and the criterion used for optimizing the embedding/detection parameters. Therefore, type-II and type-III embedding/detection techniques can be evaluated considering these three issues.

### 4.1 Postprocessing Types

There are three types of postprocessing that are employed in type-III embedder/detector designs. These are *distortion compensation*, *thresholding*, and *Gaussian mapping*.

In Ref. [2], Chen *et al.* identified the capacity achieving variant of QIM as distortion compensated QIM (DC-QIM). In DC-QIM, the quantization index modulated signal is perturbed by subtracting the  $1 - \alpha^*$  scaled version of the embedding distortion  $\mathbf{X}$ . Ramkumar, *et al.* [4] proposed a thresholding type of postprocessing where the magnitude of distortions, that can be introduced to host signal samples, are limited to  $\pm \frac{\beta}{2}$ . Hence, the type-III codeword  $\mathbf{X}_n$  is generated by limiting the values of  $\mathbf{X}$  and the processing distortion  $\mathbf{X}_t$ , in this case, is the thresholding noise. Perez-Gonzalez *et al.* [5], considering uniform scalar quantization, proposed to generate the processing distortion  $\mathbf{X}_t$  from  $\mathbf{X}$  by transforming each *iid* component  $X$  into a zero-mean Gaussian distributed random variable with a variance of  $\sigma_v^2$ . Corresponding expressions for the processing distortion  $\mathbf{X}_t$  and the codeword  $\mathbf{X}_n$  for the three types of postprocessing are as given in Table 3, where  $Q^{-1}(\cdot)$  is the inverse Gaussian Q-function.

Table 3  
Expressions for  $\mathbf{X}_t$  and  $\mathbf{X}_n$

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

#### 4.1.1 Vectorial Embedding and Detection

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

$$\sigma_X^2 = \frac{1}{N}\|\mathbf{X}\|^2 = \frac{1}{N}\left\|\frac{\mathbf{X}_n}{\alpha^*}\right\|^2 = \frac{(P + \sigma_Z^2)^2}{P} \quad (24)$$

where  $\alpha^* = \frac{P}{P + \sigma_Z^2}$ . It should be noted that, the variance of the *iid* components of the channel input  $\mathbf{X}$  (the power of the input  $\mathbf{X}$ ) in Eq. (14) is the same as the power of the optimal embedding distortion  $\mathbf{X}$  found in Eq. (24),  $\sigma_X = \sigma_X^*$ . Therefore, distortion compensation is the optimal processing when the embedding distortion is Gaussian distributed. This can be satisfied by the use of high-dimensional quantization for embedding which yields Gaussian distributed quantization error (assuming  $\mathbf{C}$  is white) [8,28]. However, a capacity achieving embedding/detection scheme based on thresholding or Gaussian mapping types of postprocessing is not possible since the relation between  $\mathbf{X}$  and  $\mathbf{X}_t$  is not linear.

#### 4.1.2 Scalar Embedding and Detection

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

#### 4.2 Forms of Demodulation

Detection of the sent message is achieved either by sample-wise hard decisions or soft decisions based on the availability of the set of watermark signals at the extractor side. The presence of watermark signals leads to an improved detection of the sent message since they can be utilized in detection operation [12,4].

There are two forms of demodulation employed in detection of the sent mes-

sage. In [12,3,5], demodulation of the sent message, from the received signal  $\mathbf{Y}$ , is realized by minimum distance decoding, and in Ref. [4], demodulation takes the form of maximum correlation rule.

#### 4.2.1 Minimum Distance Decoder

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

$$\hat{m} = \mathcal{D}(\mathbf{Y}) = \arg \min_m \|\mathbf{Y}_m - Q_\Delta(\mathbf{Y}_m)\|, \quad 1 \leq m \leq M. \quad (25)$$

It should be noted that Eq. (25) is a minimization of the quantization error over all quantizers. For the case of scalar quantization,  $Q_\Delta(\cdot)$  takes the form of dithered quantizer  $q_\Delta(\cdot)$ .

Fig. 7 displays the detectors for the binary signaling case where the embedding operation is based on scalar quantization. In the figure, the symbols  $\times$  and  $\circ$  denote the reconstruction points of the quantizers associated with the watermark sample values of  $-\frac{\Delta}{4}$  and  $\frac{\Delta}{4}$ . (However, it should be noted that, within the scope of DM, any two sample values with  $\frac{\Delta}{2}$  difference are valid choices, see Eq. (23).) When the extractor has no access to the watermark signals but only knows the reconstruction points, each sample of the embedded watermark signal is detected from each coefficient  $Y$  of the received signal  $\mathbf{Y}$  by individual hard decisions as

$$\hat{W}_i = \arg \min_{W_i \in \Omega} \|Y_i + W_i - q_\Delta(Y_i + W_i)\| \quad \text{for, } i = 1, \dots, N \quad (26)$$

where  $\Omega$  is the set of signal representations for watermark samples. Eq. (26) is based on determining the minimum Euclidean distance of the received signal coefficients to reconstruction points which can equivalently be achieved by mapping each coefficient  $Y$  over the square wave function displayed in Fig. 7-a. Then, the extracted binary watermark samples,  $\hat{W}_1, \dots, \hat{W}_N$ , are combined into the sequence  $\hat{\mathbf{W}}$  to generate the embedded watermark signal. On the other hand, when the watermark signals are present at the detector, detection of each sample is by soft decisions. Accordingly, each coefficient  $Y_m$  of  $\mathbf{Y}_m$  is mapped over the sawtooth function displayed in Fig. 7-b. The norm of the resulting signal values is the distance between  $\mathbf{Y}$  and  $\mathbf{W}_m$ . Hence, the watermark signal that has the minimum distance to  $\mathbf{Y}$  is regarded as the embedded signal.

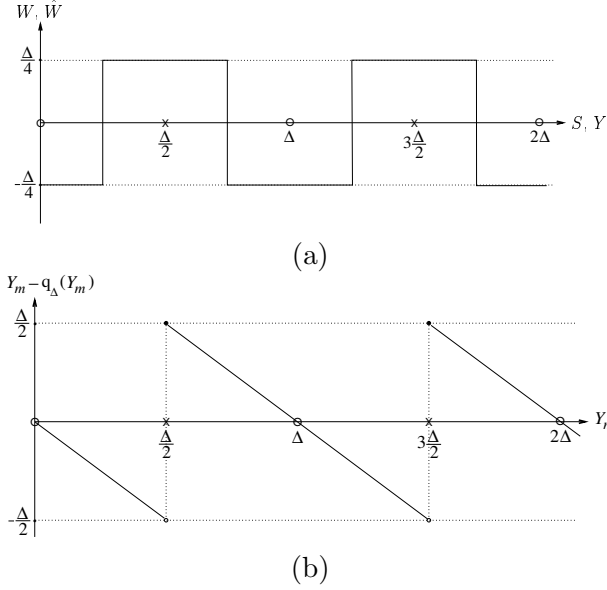


Fig. 7. Demodulation for DM based on (a) hard decisions and (b) soft decisions.

#### 4.2.2 Maximum Correlation Rule

Since the hard decisions are caused by the discontinuities in the extraction function, Fig. 7-a, an alternate approach can be taken by using a continuous extraction function, which enables extraction of the embedded watermark signal by soft decisions. When the demodulation scheme is based on maximum correlation rule, watermark signals are assumed to be present at the detector. In this form of demodulation, at first, an estimate  $\hat{\mathbf{W}}$  of embedded  $\mathbf{W}$  is extracted from the received signal. Then, the sent message is detected by matching the estimate of the embedded watermark signal to one of the watermark signals using a correlation based similarity measure as

$$\hat{\mathbf{W}} = \mathcal{D}(\mathbf{Y}), \hat{m} = \arg \max_m \frac{\mathbf{W}_m \hat{\mathbf{W}}}{\|\mathbf{W}_m\| \|\hat{\mathbf{W}}\|}, 1 \leq m \leq M. \quad (27)$$

In [4], a continuous periodic triangular extraction function is proposed. Fig. 8 displays the corresponding function used for extracting the embedded binary watermark samples that are confined to values  $-\frac{\Delta}{4}$  and  $\frac{\Delta}{4}$  for maximum separation,  $\Omega = \{-\frac{\Delta}{4}, \frac{\Delta}{4}\}$ . An estimate of the embedded watermark signal is obtained by mapping each coefficient of  $\mathbf{Y}$  over the periodic triangular function, rather than making a hard decision by the Euclidean distance decoder. As a result, each extracted sample  $\hat{W}$  is a real valued signal in the range of  $[-\frac{\Delta}{4}, \frac{\Delta}{4}]$ . Message detection is achieved by combining the sample estimates into  $\hat{\mathbf{W}} = (\hat{W}_1, \dots, \hat{W}_N)$  and then matching  $\hat{\mathbf{W}}$  to one of  $\mathbf{W}_1, \dots, \mathbf{W}_M$ .

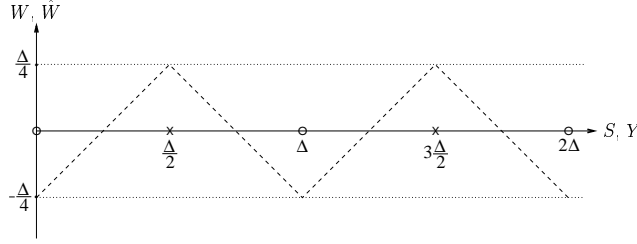


Fig. 8. Periodic extraction function.

### 4.3 Optimization Criteria for Embedding and Detection Parameters

The embedding/detection operations are controlled by a pair of parameters. The values for these parameters are optimized for the given channel noise and permitted distortion levels,  $\sigma_Z^2$  and  $P$ . One of the parameters common to all techniques is  $\Delta$  which designates the distance between the reconstruction points of the embedding quantizers. The choice of  $\Delta$  determines the embedding distortion due to quantization, and it is known to both embedder and detector. The other parameter controls the amount of processing distortion introduced to quantized signal (type-II embedded signal) by the postprocessing, and it limits the distortion due to embedding operation to the permitted amount. This parameter is known only to embedder and parameterized depending on the type of postprocessing. The values for the two interdependent parameters can be optimized based on various performance criteria as discussed in the following sections.

#### 4.3.1 Optimization of Parameters for Vectorial Embedding and Detection

Researchers in Ref. [2] optimized the embedding/detection parameters by maximizing the ratio of the embedding distortion to the sum of processing and channel distortions,  $\frac{\sigma_X^2}{\sigma_{X_t}^2 + \sigma_Z^2}$ , at the extractor as

$$(\Delta, \sigma_{X_t}^2) = \arg \max_{\Delta, \sigma_{X_t}^2} \left\{ \frac{\sigma_X^2}{\sigma_{X_t}^2 + \sigma_Z^2} \left| \sigma_Z^2, \frac{1}{N} \|\mathbf{X}_n\|^2 = P, \mathbf{X}_t \right. \right\}. \quad (28)$$

With the use of high dimensional quantization for embedding/detection, the marginal *pdf* of embedding distortion  $\mathbf{X}$  approximates Gaussian distribution and consequently distortion compensation becomes the optimal processing. Hence, for the given channel noise level,  $\Delta$  and  $\alpha$  are selected such a way that Eq. (28) is satisfied (by maximizing over  $\alpha$ ), where  $\mathbf{X}_t = (1 - \alpha)\mathbf{X}$  and  $\mathbf{X}_n = \alpha\mathbf{X}$ , *i.e.*  $\sigma_{X_t}^2 = (1 - \alpha)^2\sigma_X^2$  and  $\sigma_X^2 = \frac{P}{\alpha^2}$ . This leads to  $\alpha = \frac{P}{P + \sigma_Z^2}$  which is in accord with the results of Section 2.1 due to the duality between the two channel models.

### 4.3.2 Optimization of Parameters for Scalar Embedding and Detection

Researchers in [4,3,5] modeled the effective noise that distorts the embedded watermark signal in terms of the statistics of the channel noise  $\mathbf{Z}$  and the processing distortion  $\mathbf{X}_t$ ,  $\mathbf{Z}_{eff} = \mathbf{Z} - \mathbf{X}_t$ . The optimum values for embedding/detection parameters are then selected in a way that the distortion in the extracted watermark signal is minimized. Ref. [3] obtained the  $\alpha$  values for distortion compensation type of postprocessing, rather than assuming  $\alpha^* = \frac{P}{P+\sigma_Z^2}$ , and provided the approximation  $\alpha = \sqrt{\frac{P}{P+2.71\sigma_Z^2}}$ . Expressions for the optimal values of  $\Delta$  and  $\beta$  for the thresholding type of postprocessing are reported in [4]. Although Ref. [5] does not provide the  $\sigma_v$  values for Gaussian mapping, the derivation procedure is straightforward.

When the host signal is uniformly distributed over all quantization intervals, the embedding distortion  $X$  introduced to each host signal coefficient  $C$  is uniformly distributed in  $[-\frac{\Delta}{2}, \frac{\Delta}{2}]$ . Given that the host signal is *iid*,  $\mathbf{X}$  and  $\mathbf{X}_t$  are *iid* random vectors with the marginal distributions given above, since the embedding operation is memoryless. Correspondingly, the *pdf* and statistics of processing distortion  $X_t$  and the codeword  $X_n$  can be determined for a given type of postprocessing in terms of the step size  $\Delta$  and the threshold  $\beta$  or the scaling factor  $\alpha$  or the variance  $\sigma_v^2$ . Table 4 parameterizes the *pdf* of processing distortion  $X_t$  assuming uniformly distributed embedding distortion  $X$ . It should also be noted that, for large  $N$ , the distortion  $P$  introduced to host signal  $\mathbf{C}$ , due to embedding operation, is equal to  $\sigma_{X_n}^2$ , *i.e.*  $\frac{1}{N} \|\mathbf{X}_n\|^2 = P \rightarrow \sigma_{X_n}^2$  as  $N \rightarrow \infty$ .

Table 4  
 $f_{X_t}(x_t)$  for the three types of postprocessing

Postprocessing	$f_{X_t}(x_t)$
Thresholding	$f_{X_t}(x_t) = \frac{\beta}{\Delta} \delta(x_t) + \frac{1}{\Delta} \text{rect}(\Delta - \beta)$
Distortion Compensation	$f_{X_t}(x_t) = \frac{1}{(1-\alpha)\Delta} \text{rect}((1-\alpha)\Delta)$
Gaussian mapping	$\frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left(-\frac{z_{eff}^2}{2\sigma_v^2}\right)$

Assuming  $Z$  and  $X_t$  are independent, the resulting *pdf* of  $Z_{eff}$ ,  $f_{Z_{eff}}(z_{eff})$  can be computed by the convolution of the individual *pdfs*  $f_Z(z)$  and  $f_{X_t}(x_t)$ . Thus, for  $Z \sim \mathcal{N}(0, \sigma_Z^2)$   $f_{Z_{eff}}(z_{eff})$  corresponding to different types of postprocessing can be derived as given in Table 5, where  $\text{erf}(\cdot)$  is the Gaussian error function,  $\text{erf}(z) = \frac{2}{\pi} \int_0^z e^{-x^2} dx$ .

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

*Maximizing Correlation:* With this criterion, the selection of parameters is based on maximizing the normalized-correlation between the embedded and

Table 5

$f_{Z_{eff}}(z_{eff})$  for the three types of postprocessing

Postprocessing	$f_{Z_{eff}}(z_{eff})$
Thresholding	$\frac{\beta}{\Delta\sqrt{2\pi\sigma_Z^2}} e^{-\frac{z_{eff}^2}{2\sigma_Z^2}} + \frac{1}{2\Delta} \left( \operatorname{erf}\left(\frac{z_{eff} + \frac{\Delta-\beta}{2}}{\sqrt{2}\sigma_Z}\right) - \operatorname{erf}\left(\frac{z_{eff} - \frac{\Delta-\beta}{2}}{\sqrt{2}\sigma_Z}\right) \right)$
Distortion Compensation	$\frac{1}{2(1-\alpha)\Delta} \left( \operatorname{erf}\left(\frac{z_{eff} + \frac{(1-\alpha)\Delta}{2}}{\sqrt{2}\sigma_Z}\right) - \operatorname{erf}\left(\frac{z_{eff} - \frac{(1-\alpha)\Delta}{2}}{\sqrt{2}\sigma_Z}\right) \right)$
Gaussian mapping	$\frac{1}{\sqrt{2\pi(\sigma_Z^2 + \sigma_v^2)}} \exp\left(-\frac{z_{eff}^2}{2(\sigma_Z^2 + \sigma_v^2)}\right)$

extracted watermark signals [4]. Since  $\mathbf{Z}_{eff}$  is the noise that distorts the type-II codeword  $\mathbf{X}$  corresponding to watermark signal  $\mathbf{W}$ , the signal  $\hat{\mathbf{W}}$  extracted from  $\mathbf{Y}$  can be expressed in terms of  $\mathbf{Z}_{eff}$  and  $\mathbf{W}$  using the extraction function shown in Fig. 8. (Note that if  $\mathbf{Z}_{eff} = 0$ , then  $\mathbf{W} = \hat{\mathbf{W}}$ .) Hence, a binary distributed watermark signal sample  $W$  with values in  $\{-\frac{\Delta}{4}, \frac{\Delta}{4}\}$  embedded in a host signal coefficient is extracted as

$$\hat{W} = \begin{cases} \left(\frac{(2i+1)\Delta}{4} - Z_{eff}\right)(-1)^i, & i\frac{\Delta}{2} < Z_{eff} \leq \frac{(i+1)\Delta}{2}, i \in \mathcal{Z} \text{ if } W = \frac{\Delta}{4}, \\ \left(-\frac{(2i+1)\Delta}{4} + Z_{eff}\right)(-1)^i, & i\frac{\Delta}{2} < Z_{eff} \leq \frac{(i+1)\Delta}{2}, i \in \mathcal{Z} \text{ if } W = -\frac{\Delta}{4}. \end{cases} \quad (29)$$

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

$$\rho = \frac{\mathbf{W}^T \hat{\mathbf{W}}}{\|\mathbf{W}\| \|\hat{\mathbf{W}}\|} = \frac{E[W\hat{W}]}{\sqrt{E[W^2]E[\hat{W}^2]}} = \frac{R(1)}{\sqrt{R(2)}}, \quad (30)$$

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

$$R(p) = 2 \sum_{i=0}^{i=\infty} \int_{\frac{i\Delta}{2}}^{\frac{(i+1)\Delta}{2}} \left( \left( \frac{(2i+1)\Delta}{4} - z_{eff} \right) (-1)^i \right)^p f_{Z_{eff}}(z_{eff}) dz_{eff}. \quad (31)$$

Therefore, the optimal parameter values for the utilized postprocessing technique is computed by maximizing Eq. (30) over  $\Delta$  and  $\sigma_{X_t}^2$  using the *pdfs* given in Table 5 for the given channel noise level and permitted distortion as in Eq. (28).

*Minimizing Probability of Error:* In a similar manner, the embedding/detection parameters can be selected to minimize the probability of error in detecting an embedded watermark sample [5]. Since  $Z_{eff}$  indicates the deviation of the

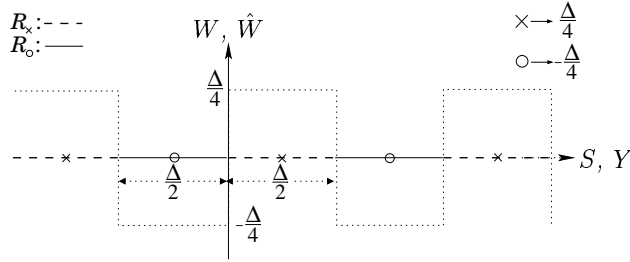


Fig. 9. Embedding and detection of a binary watermark sample.

received signal coefficient  $Y$  from the reconstruction points, the probability of detection error,  $P_e$ , can be calculated by integrating  $f_{Z_{eff}}(z_{eff})$  over all decision regions but excluding the one associated with the sent sample. For the binary signaling case depicted in Fig. 9, the symbols  $\times$  and  $\circ$  denote the reconstruction points of two quantizers associated with sample values  $\frac{\Delta}{4}$  and  $-\frac{\Delta}{4}$ , respectively. The decision regions  $R_\times$  and  $R_\circ$  are used to map the received signal coefficient  $Y$  to  $\frac{\Delta}{4}$  or  $-\frac{\Delta}{4}$  by hard decisions. Assuming  $\frac{\Delta}{4}$  and  $-\frac{\Delta}{4}$  are equally likely to be embedded, corresponding  $P_e$  is calculated as

$$P_e = \Pr\{Y \in \mathcal{R}_\circ \mid w = \frac{\Delta}{4}\} = \int_{\mathcal{R}_\circ} f_{Z_{eff}}(z_{eff} - \frac{\Delta}{4}) dz_{eff}. \quad (32)$$

Then, the parameters can be selected to minimize  $P_e$  for the given  $P$ ,  $\sigma_Z^2$ , and the type of postprocessing.

*Maximizing Mutual Information:* Alternately, the parameters can be selected to maximize the mutual information between the embedded watermark sample  $W$  and the received signal coefficient  $Y$  [3]. The mutual information between  $W$  and  $Y$  is expressed as

$$I(W, Y) = H(Y) - H(Y|W). \quad (33)$$

As the erroneous detection of  $W$  from  $Y$  is due to the noise  $Z_{eff}$ ,  $H(Y|W)$  in Eq. (33) can be computed in terms of the effective noise *pdf* conditioned on  $W$ ,  $f_{z_{eff}|w}(z_{eff}|w)$ . The *pdf*  $f_{Z_{eff}|W}(z_{eff}|w)$  can be calculated over any quantization interval  $\Delta$ , since the signal constellation is periodic with  $\Delta$  (reconstruction points corresponding to quantizer associated with  $W$  are  $\Delta$  apart). However, one should take into account that when  $Z_{eff}$  is heavy tailed (the range of  $f_{Z_{eff}}(z_{eff})$  is larger than  $\Delta$ ), its *pdf* will be wrapped around  $\Delta$  due to the periodicity. Consequently,  $H(Y)$  is computed from  $H(Y|W)$  by averaging it over  $W$ . With this criterion, optimization of parameter values is by maximizing Eq. (33) for the given distortion constraints over  $\Delta$  and  $\sigma_{X_t}^2$ .

## 5 Performance Comparisons

Fig. 10 displays the achievable data hiding rates of various embedding/detection techniques for the binary signaling case, obtained using Eqs. (3) and (33), compared to hiding rates of type-I (additive scheme) and optimal type-III (capacity). The hiding rate is measured in the number of bits that can be hidden into a host signal coefficient, and the robustness measure is defined in terms of the ratio between the embedding distortion power and the channel noise power,  $WNR = 10 \log_{10} \frac{P}{\sigma_z^2}$  in dB. However, for type-I methods, WNR by itself cannot be the indicator of the robustness as the host signal is considered to be a part of the noise. Therefore, another measure that can be considered is the ratio of the host signal power to embedding distortion power,  $DWR = 10 \log_{10} \frac{\sigma_c^2}{P}$  in dB. The embedding/detection parameters for type-II and type-III methods are selected so that the hiding rate is maximized, Eq. (33). The additive scheme (type-I) and DM (type-II) have preferable performances, respectively, at very low and very high WNRs. For DM, the gap with the upper bound at higher WNRs is due to binary signaling. Thus, the performance can be improved for multi-level signal representations. At high WNRs, additive scheme has a fixed hiding rate that is well below the capacity. On the other hand, hiding rate of DM drops exponentially with the decreasing WNR. The poor performance of both methods in mid-WNR range is due to non-optimal codebook designs.

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

Type-III methods employing thresholding and distortion compensation types of postprocessing perform closely in the whole WNR range. On the other hand, Gaussian mapping processing has a comparable performance only for WNRs higher than  $-7.8$  dB. Below that range the rate drops rapidly. At WNRs lower than  $-8.7$  dB thresholding performs marginally better, while from  $-8.7$  dB to  $-7$  dB, distortion compensation performs best. Above  $-7$  dB, both distortion compensation and Gaussian mapping are the preferred postprocessing types. Fig. 11 shows the hiding rates for the corresponding methods with multi-level signaling. With the decreasing noise level and higher signal representation levels, all methods yield similar data hiding rates as the need for postprocessing reduces. Ultimately when there’s no noise, the DM is the optimal embedding/detection technique.

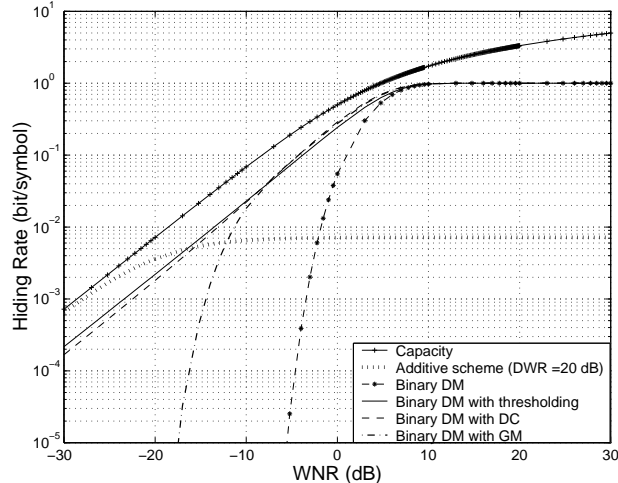


Fig. 10. Comparison of the hiding rates corresponding to various hiding methods considering binary signaling obtained for  $P = 10$ .

The normalized-correlation,  $\rho$ , and probability of error,  $P_e$ , performances for the considered methods are respectively given in Figs. 12-a and 12-b. The corresponding embedding/detection parameters for the hiding methods are selected as described in Sec. 4.3.2. The correlation between an embedded binary watermark signal  $\mathbf{W}$  and extracted watermark signal  $\hat{\mathbf{W}}$  is calculated by using Eq. (30), and the probability of the error in detecting an embedded binary watermark sample is computed by using Eq. (32).

The relative performances of the three types of postprocessing obtained for the three criteria, Figs. 10, 12-a, and 12-b are in accord with each other. Thus, thresholding type of postprocessing performs better when WNR is below approximately  $-9$  dB, and at higher WNRs distortion compensation has better performance. Above  $-7$  dB, Gaussian mapping and distortion compensation have comparable performances.

One intuitive way to evaluate the performance characteristics of type-I, type-II, and type-III methods at varying noise levels is by considering the size of decision cells at the detector. For type-II methods in the absence of noise, the extracted watermark signals correspond to reconstruction points of the embedding quantizers. Thus, decision cells can collapse to points and the data hider can afford to use higher level signaling without any performance penalty. However, with the increasing noise level, the successful extraction of the embedded watermark signal requires decision cells to be enlarged accordingly. In type-III methods  $\Delta$  is increased in accordance with the channel noise level  $\sigma_Z^2$  and the corresponding increase in embedding distortion due to increased  $\Delta$  is compensated by the postprocessing. Hence, the data hider has the freedom to change the size of the decision cell depending on the noise level. Ultimately when the noise level is very high, the optimal strategy becomes making the decision regions arbitrarily large as in type-I methods where even for very

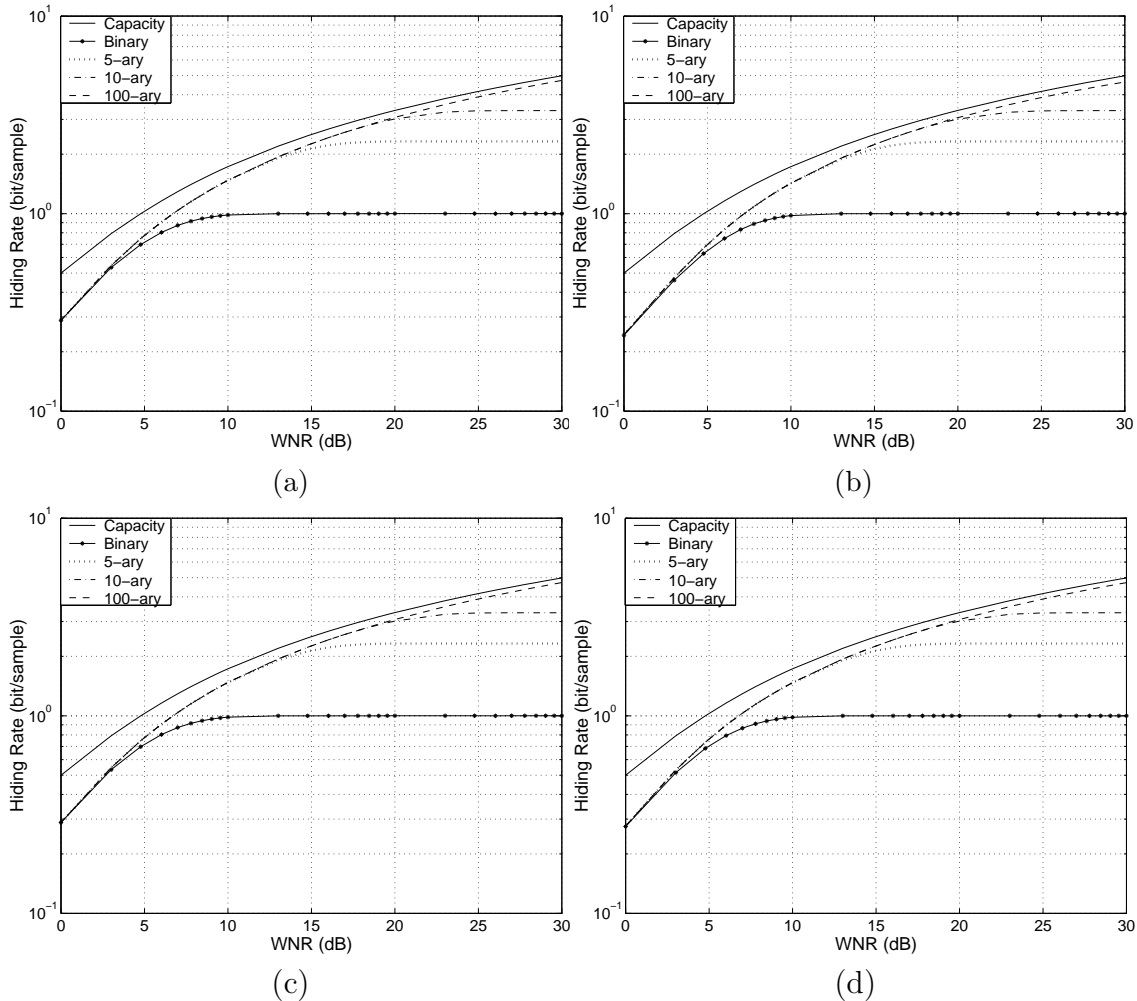


Fig. 11. Data hiding rates for (a) DM and DM followed by (a) thresholding, (b) distortion compensation, (c) Gaussian mapping types of postprocessing with with binary, 5-ary, 10-ary, and 100-ary signaling.

high noise levels, the detector is able to find some traces of the embedded watermark signals.

## 6 Conclusions

In this paper, we introduce the CAE-CID framework, that is equivalent to framework introduced by Costa, with a data hiding perspective. Within this approach, the codeword generation at the encoder depends on the design  $\mathbf{U} = \mathbf{X} + \mathbf{C}$  and the constraint  $\frac{1}{N}\|\mathbf{X} - \mathbf{X}_t\|^2 \leq P$  rather than  $\mathbf{U} = \mathbf{X} + \alpha\mathbf{C}$  and  $\frac{1}{N}\|\mathbf{X}\|^2 \leq P$  where the statistics of  $\mathbf{X}_t$ , and  $\alpha$  are channel dependent. In the CAE-CID framework, the encoding operation is formulated in terms of channel input  $\mathbf{X}_n = \mathbf{X} - \mathbf{X}_t$ , and the decoding operation does not require channel noise information. Since  $\mathbf{X}$  and  $\mathbf{X}_t$  can be dually interpreted as the em-

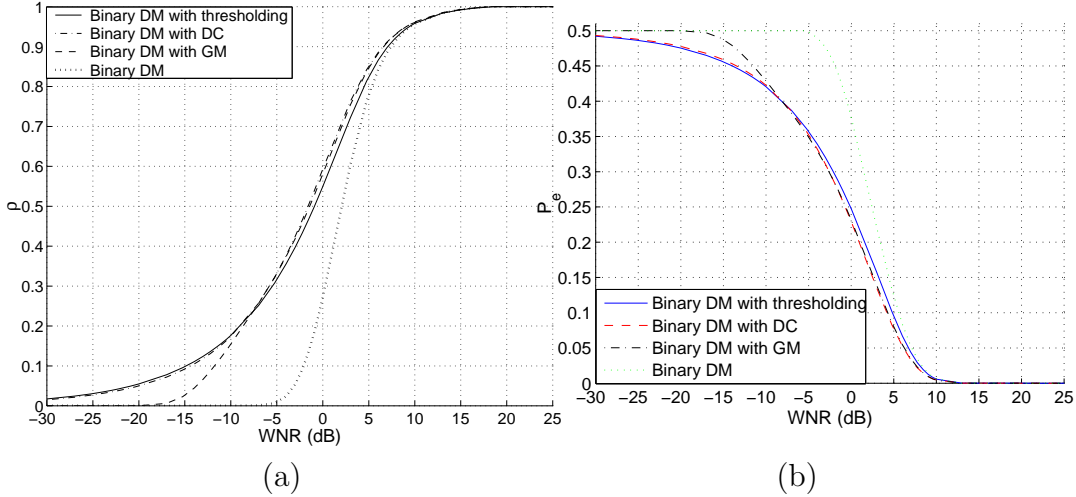


Fig. 12. (a) The normalized-correlation between  $\mathbf{W}$  and  $\hat{\mathbf{W}}$  and (b) the probability of error in detecting  $W$  for the considered hiding methods when  $P = 10$ .

bedding distortion and processing distortion, respectively, within the context of data hiding, practical embedder/detector designs that employ quantizers in embedding followed by a processing, like thresholding, distortion compensation, or Gaussian mapping fit well into this framework. For AWGN attack and mean squared error distortion measure, results indicate that distortion compensation is the optimal postprocessing if  $\mathbf{X}$ ,  $\mathbf{X}_t$ , and  $\mathbf{C}$  are Gaussian distributed. However, for uniform distribution of  $\mathbf{X}$  the optimal processing depends on the channel noise level and the dependence between  $\mathbf{X}$  and  $\mathbf{X}_t$ .

We identify the key characteristics of quantization-based embedding/detection techniques that enable a proper evaluation of such data hiding methods. These characteristics are the type of postprocessing, the form of demodulation, and the optimization criteria for the embedding/detection parameters. Performance comparison of the embedding/detection techniques based on probability of error, correlation, and mutual information metrics lead to the same conclusion. For the two extreme cases of “severe noise” and “low noise,” additive schemes and dither modulation (no postprocessing) achieve the optimal performance, respectively. However, for all other noise levels, the two schemes do not have preferable performances. At relatively high noise levels, embedding/detection with thresholding type of processing performs best. While distortion compensation performs closely to thresholding, Gaussian mapping is not suited for high noise level applications. For low noise levels, on the other hand, both distortion compensation and Gaussian mapping types of processing yield comparable performances.

## References

- [1] M. Costa, Writing on dirty paper, *IEEE Transactions on Information Theory* 29 (1983) 439–441.
- [2] B. Chen, G. Wornell, Preprocessed and postprocessed quantization index modulation methods for digital watermarking, in: *Proc SPIE: Security and Watermarking of Multimedia Contents II*, Vol. 3971, 2000, pp. 48–59.
- [3] J. J. Eggers, J. K. Su, B. Girod, A blind watermarking scheme based on structured codebooks, *IEE Colloq. Secure Images and Image Authentication 4* (2000) 1–6.
- [4] M. Ramkumar, A. N. Akansu, Self-noise suppression schemes for blind image steganography, in: *Proc SPIE International Workshop on Voice, Video and Data Communication, Multimedia Applications*, Vol. 3845, 1999.
- [5] F. Perez-Gonzalez, F. Balado, J. R. Hernandez Martin, Performance analysis of existing and new methods for data hiding with known-host information in additive channels, *IEEE Transactions on Signal Processing* 51 (4) (2003) 960–980.
- [6] I. J. Cox, M. L. Miller, A. L. McKellips, Watermarking as communication with side information, *Proc. of IEEE* 87 (1999) 1127–1141.
- [7] J. Chou, S. S. Pradhan, L. E. Ghaoui, K. Ramchandran, On the duality between data hiding and distributed source coding, in: *Proc. of 33rd Annual Asilomar conference on Signals, Systems, and Computers*, 1999.
- [8] R. J. Barron, B. Chen, G. W. Wornell, The duality between information embedding source coding with side information and its implications—applications, *IEEE Transactions on Information Theory* 49 (5) (2003) 1159–1180.
- [9] B. Chen, G. W. Wornell, Quantization index modulation: A class of provably good methods for digital watermarking and information embedding, *IEEE Transactions on Information Theory* 47 (4) (2001) 1423–1443.
- [10] J. Chou, S. S. Pradhan, L. E. Ghaoui, K. Ramchandran, A robust optimization solution to the data hiding problem using distributed source coding principles, in: *Proc SPIE: Image and Video Communications and Processing*, Vol. 3974, 2000.
- [11] R. Zamir, S. Shamai, U. Erez, Nested linear/lattice codes for structured multiterminal binning, *IEEE Transactions on Information Theory* 48 (5) (2002) 1250–1276.
- [12] B. Chen, G. W. Wornell, Dither modulation: A new approach to digital watermarking and information embedding, in: *Proc. of SPIE: Security and Watermarking of Multimedia Contents*, Vol. 3657, 1999, pp. 342–353.

- [13] K. Tanaka, Y. Nakamura, K. Matsui, Embedding secret information into a dithered multi-level image, in: Proc. of IEEE International Conference On Image Processing, 1990, pp. 216–220.
- [14] R. G. van Schyndel, A. Z. Tirkel, C. F. Osborne, A digital watermark, in: Proc. of IEEE International Conference On Image Processing, Vol. 2, 1994, pp. 86–90.
- [15] G. Caronni, Assuring ownership rights for digital images, in: Proc. of Reliable IT Systems, Vol. VIS-95, Vieweg Publishing Company, 1995.
- [16] M. D. Swanson, B. Zhu, A. H. Tewfik, Data hiding for video-in-demand, in: Proc. of IEEE International Conference On Image Processing, Vol. 2, 1997, pp. 676–679.
- [17] H.-J. M. Wang, P.-C. Su, C.-C. J. Kuo, Wavelet-based digital image watermarking, *Optics Express* 3 (12) (1998) 491–496.
- [18] M. Wu, B. Liu, Watermarking for image authentication, in: Proc. of IEEE International Conference On Image Processing, Vol. 2, 1998, pp. 437–441.
- [19] J. J. Eggers, R. Bauml, R. Tzschoppe, B. Girod, Scalar Costa scheme for information embedding, *IEEE Transactions on Signal Processing* 51 (4) (2003) 1003–1019.
- [20] S. I. Gelfand, M. S. Pinsker, Coding for channel with random parameters, *Problems of Control and Information Theory* 9 (1) (1980) 19–31.
- [21] P. Moulin, J. A. O’Sullivan, Information-theoretic analysis of information hiding, *IEEE Transactions on Information Theory* 49 (2003) 563–593.
- [22] A. S. Cohen, A. Lapidoth, The gaussian watermarking game, *IEEE Transactions on Information Theory* 48 (2002) 1639–1667.
- [23] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, Second Edition, New York: John-Wiley & Sons Inc., 1991.
- [24] U. Erez, S. Shamai, R. Zamir, Capacity and lattice-strategies for cancelling known interference, in: Proc. of Int. Symp. Information Theory and Its Applications, 2000, pp. 681–684.
- [25] M. Ramkumar, Data hiding in multimedia - theory and applications, Ph.D. thesis, New Jersey Institute of Technology, Newark, NJ (2000).
- [26] H. T. Sencar, M. Ramkumar, A. N. Akansu, Efficient codebook structures for practical information hiding systems, in: Proc. of CISS, 2001.
- [27] R. G. Gray, T. M. Stockham, Dithered quantizers, *IEEE Transactions on Information Theory* 39 (3) (1993) 805–812.
- [28] R. M. Gray, D. L. Neuhoff, Quantization, *IEEE Transactions on Information Theory* 44 (6) (1998) 2325–2383.