Informed Watermarking by Means of Orthogonal and Quasi-Orthogonal Dirty Paper Coding

Andrea Abrardo, *Member, IEEE,* and Mauro Barni, *Member, IEEE*

***Abstract*—A new dirty paper coding technique that is robust against the gain attack is presented. Such a robustness is obtained by adopting a set of (orthogonal) equi-energetic codewords and a correlation-based decoder. Due to the simple structure of orthogonal codes, we developed a simple yet powerful technique to embed the hidden message within the host signal. The proposed technique is an optimal one, in that the embedding distortion is minimized for a given robustness level, where robustness is measured through the maximum pairwise error probability in the presence of an additive Gaussian attack of given strength. The performance of the dirty coding algorithm is further improved by replacing orthogonal with quasi- orthogonal codes, namely, Gold sequences, and by concatenating them with an outer turbo code. To this aim, the inner decoder is modified to produce a soft estimate of the embedded message. Performance analysis is carried out by means of extensive simulations proving the validity of the novel watermarking scheme.**

***Index Terms*—Dirty paper coding, informed embedding, informed watermarking, orthogonal codes, spherical codes, turbo coding.**

I. Introduction

SINCE it has been recognized that digital watermarking can be seen as a problem of communications with side information at the encoder [1], informed watermarking algorithms, which adapt the watermarking signal to the particular realization of the to-be-marked signal, have received increasing attention due to their superior performance with respect to conventional blind embedding methods. By recalling some consolidated results of Inf. Theory dealing with channels with side information [2], [3], and by adapting them to the watermarking case [4]–[7], it can be demonstrated that under the hypothesis of additive white noise attack and independent and identically distributed (iid) Gaussian host features [7], it is possible to completely eliminate the interference between the hidden and the host signals, thus reaching the same capacity obtained by systems where the decoder can access the original non marked data. An extension of the above result to the more general cases of non-iid sequences and non-Gaussian host features is given in [8], [9].

Manuscript received June 28, 2003; revised September 16, 2004. This work was supported in part by the Italian Ministery for Instruction, University, and Research (MIUR) under project WAtermarking for Video and imagEs COpyright Protection (WAVECOP) and by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT. The information in this paper is provided as is, and no guarantee or warranty is given or implied that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. The associate editor coordinating the review of this paper and approving it for publication was Prof. Pierre Moulin.

The authors are with the Department of Information Engineering, University of Siena, 53100 Siena, Italy (e-mail: abrardo@dii.unisi.it; barni@dii.unisi.it).

Digital Object Identifier 10.1109/TSP.2004.839909

Quantization Index Modulation (QIM) watermarking is a practical way to put the informed watermarking principles at work. A scalar version of QIM, known as Dither Modulation (DM) with bit repetition, or its distortion compensated version (DC-DM or SCS [10]), is usually adopted, since, due to the lattice-based nature of the codebook entries, it permits very fast coding and decoding, with very good performance. Alternatively, the correlation between the cover feature sequence and a reference pseudo-noise sequence is quantized, leading to the Spread Transform-Dither Modulation (ST-DM) algorithm [6], [11].

A major problem with lattice-based watermarking is vulnerability against value-metric scaling of the host features, specifically against the gain attack, consisting in the multiplication of the host feature sequence by a constant factor g, which is unknown to the decoder. This weakness derives from the choice of using the host features amplitude to convey the hidden message. The possible approaches to cope with the gain attack, in the framework of QIM watermarking, can be classified into three main categories: i) embedding of an auxiliary pilot signal to be used by the decoder to recover from amplitude scaling [12][1]; ii) embedding the watermark in a domain which is invariant to value-metric scaling; and iii) adoption of spherical codewords [13] together with correlation decoding [14], [15], [19].

The insertion of a pilot signal is not advisable in many practical applications, since embedding two signals instead of one increases the distortion incurred by the host signal, and most of all, it introduces an additional source of weakness against malicious attacks, since attackers can either decide to attack the watermark or the pilot signal. The identification of an embedding domain that is invariant with respect to the gain g could represent the ideal solution; however, finding such a domain is not an easy task.

The third possibility, which is the one we follow here, stems from the observation that using a minimum distance decoder on a set of codewords lying on the surface of a sphere, i.e., adopting equi-energetic codewords, results in hyperconic decoding regions centered in the origin. This in turn ensures that multiplication by a constant factor does not move a point from one decoding region to the other. The adoption of spherical codes in watermarking can also be given a more theoretical justification if we assume that the host features follow a Gaussian distribution. The analysis given in [7] shows that distributing the codewords on the surface of a sphere and using a minimum distance decoder results in a capacity achieving strategy. It is worth noting, though, that invariance against the gain attack is obtained even in the more general case of non-Gaussian host

[1]In principle g could also be estimated blindly; however, in this case, the accuracy of the estimate gets worse.

features. What is lost with sources that cannot be assumed to lie on a hypersphere is embedding efficiency, i.e., a higher distortion is required for the same level of robustness.

The design of good spherical codes has long been studied [13], and several classes of spherical codes exhibiting good compaction and/or error correction properties exist (see, for example, the papers by Hamkins and Zeger [16], [17]); however, in most cases, the structure of these codes is so complex that a simple, yet effective, embedding strategy is not easy to find. In fact, in practical situations, where due to limited length, the host feature sequence cannot be assumed to lie on a spherical surface, the simple embedding strategy indicated by theory [3], [7] and described in (3) cannot be used, since the host feature sequence is likely to be very distant from the closest codebook entry associated to the to-be-hidden message. In this case, a more sophisticated embedding strategy is needed to ensure that the watermarked feature sequence falls inside the correct decoding region. This is, in the essence, the ultimate goal of this paper: to propose a watermarking scheme based on spherical codes for which we developed a simple embedding strategy, which minimizes the embedding distortion under a robustness constraint. To do so, we rely on the properties of orthogonal equi-energetic codes. Then, by recognizing that orthogonal codes have rather poor error correcting capabilities, we introduce two modifications to the basic scheme that permit to improve the performance of the system: First, quasi-orthogonal, Gold sequences are used instead of orthogonal sequences; second, a channel turbo coding step is applied on top of the orthogonal codes.

The use of equi-energetic codewords and correlation-based decoding as a possible remedy against the gain attack was already proposed by Miller *et al.* in [14] and [15]. Miller's system relies on a dirty paper trellis in which several paths are associated to the same message. Though ensuring excellent performance, the system by Miller *et al.* suffers from several problems, mostly deriving from the difficulty of ensuring that the correct decoding region is entered with minimum distortion. This causes Miller's system to be very complex and the watermark to be slightly perceptible. Another scheme somewhat resembling our system is the Improved Spread Spectrum (ISS) watermarking algorithm proposed by Malvar and Florencio [18], where two equi-energetic codewords are used to embed one bit of information. A simple informed embedding strategy is also described to map the host feature sequence into the correct decoding region, while paying attention to keep the embedding distortion as low as possible. The main difference between ISS and our scheme is that ISS does not allow the encoder to choose the codewords according to the host signal, i.e., it does not follow the dirty paper coding strategy. This results in a performance penalty that is particularly evident for moderate to high rates (see Sections III-B and V-A).

Throughout the paper, the performance of the newly proposed scheme is evaluated by means of extensive numerical simulations and compared with competing gain-invariant strategies such as ISS and dirty-trellis watermarking. As demonstrated in Section V-A, a good robustness is obtained, while retaining a sufficiently high payload and keeping the computational burden extremely low.

This paper is organized as follows. In Section II, we describe the theoretical framework and notation. In Section III, the basic orthogonal dirty paper coding algorithm is presented. In Section IV, the performance of the basic algorithm are improved by adopting quasi-orthogonal instead of orthogonal codewords. The possibility of further improving the system by concatenating the dirty paper code with an outer turbo code is described in Section V. Throughout the paper, results obtained by means of Monte Carlo simulations are used to evaluate the performance of the novel algorithm. Finally, in Section VI, some conclusions are drawn and directions for future research highlighted.

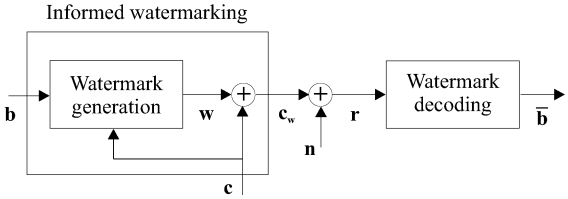


Fig. 1. General informed watermarking scheme.

II. Theoretical Framework

The general formulation of the watermarking problem as a communication channel with side information at the encoder is summarized in Fig. 1. At the input of the system, we have a message $\mathbf{b}$ chosen within the set $B = \{\mathbf{b}_i\}$ with all the possible messages. Let 2^{nR} be the number of messages contained in B[2]. The message $\mathbf{b}$ has to be embedded within a cover feature sequence $\mathbf{c}$ emitted by a source C, producing the watermarked feature sequence $\mathbf{c_w}$. The insertion of $\mathbf{b}$ within $\mathbf{c}$ goes through the definition of a watermarking signal $\mathbf{w}$, which is added to $\mathbf{c}$, hence, giving

$$\mathbf{c_w} = \mathbf{c} + \mathbf{w}. \tag{1}$$

Note that, in general, $\mathbf{w}$ may depend on $\mathbf{c}$, as it is actually dictated by the informed watermarking paradigm. The watermarked sequence is usually corrupted by an attack noise $\mathbf{n}$ that we will assume does not depend on $\mathbf{c_w}$. The decoder receives a corrupted version $\mathbf{r}$ of the marked sequence, and it produces an estimate of the hidden message.

In [3], Costa showed that under the assumptions that $\mathbf{c}$ and $\mathbf{n}$ are iid Gaussian sequences, the capacity of the channel depicted in Fig. 1 does not depend on the power of the host sequence $\mathbf{c}$. Some recent generalizations of Costa's results can be found in [7]–[9].

The capacity achieving embedding strategy following from Costa's work, which is usually referred to as random binning, proceeds as follows. The embedder generates a codebook $\mathcal{U}$ consisting of 2^{nR_t} sequences $\mathbf{u}_i$ having Gaussian distribution with zero mean and variance $\sigma_w^2 + (\alpha)^2\sigma_c^2$, where σ_c^2 indicates the variance of the cover features, σ_w^2 is the maximum power the encoder may use, and α is a parameter to be optimized based on the particular values assumed by σ_w^2 and σ_n^2. These sequences are split into 2^{nR} bins, where each bin is associated with a message in B. To cast $\mathbf{b}$ into $\mathbf{c}$, the embedder looks for a sequence $\mathbf{u}$ in the bin indexed by $\mathbf{b}$ such that

$$|(\mathbf{u} - \alpha\mathbf{c}) \cdot \mathbf{c}| < \varepsilon \tag{2}$$

[2]We measure the rate in bits/sample instead of nats/sample as in [3].

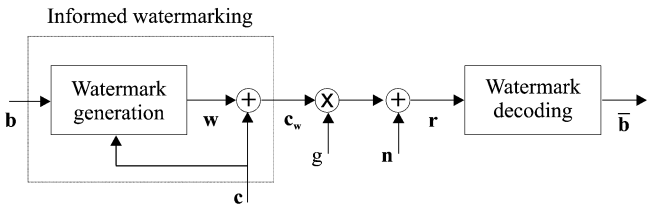


Fig. 2. General informed watermarking scheme with the gain attack.

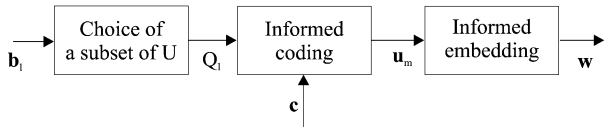


Fig. 3. Splitting of the informed watermarking procedure into informed coding and informed embedding.

for a small ε. Then, the marked feature vector $\mathbf{c_w}$ is formed as

$$\mathbf{c_w} = \mathbf{u} + (1-\alpha)\mathbf{c}. \tag{3}$$

We can observe that (2) ensures that $\mathbf{w}$ and $\mathbf{c}$ are nearly orthogonal: a condition that is equivalent to ensuring the joint typicality of $\mathbf{u}, \mathbf{c}$, and $\mathbf{c_w}$ (for a tutorial introduction to typical sequences, see [19] and [20]). Upon receiving $\mathbf{r}$, the decoder looks for a sequence in $\mathcal{U}$ that is jointly typical with $\mathbf{r}$ and outputs the message associated with the bin to which the decoded sequence belongs.

When the gain attack is taken into account, the scheme in Fig. 1 must be modified as in Fig. 2, where prior to noise addition the watermarked features are scaled by a factor g. As a matter of fact, in Costa's formulation, random binning watermarking does not imply any weakness against the gain attack. For n sufficiently large, in fact, all the typical sequences $\mathbf{u}$ have approximately the same energy, since they are uniformly distributed in a thin n-dimensional spherical shell of radius $\sqrt{\sigma_n^2 + (\alpha^*)^2\sigma_c^2}$.

In the following, we find it convenient to look at informed watermarking as a process consisting of two main steps: information coding and information embedding (see Fig. 3). Specifically, the strategy of associating more than one codeword to each message and letting the chosen codeword depend on the side information will be referred to as *informed coding*, or dirty paper coding, in honor of Costa's paper, whereas the term *informed embedding* will denote the strategy used to actually embed the chosen codeword within the host feature sequence. Note that as long as the embedder can ensure that $\mathbf{c_w}$ lies in the correct decoding region, host-signal interference is cancelled, and a zero error probability is reached in the absence of attacks. Accordingly, by relying on the simple structure of orthogonal codewords, we propose an optimal informed embedding strategy, ensuring that a given level of robustness is obtained while minimizing the embedding distortion $\|\mathbf{w}\|^2$. This is the aim of the next section, where a robustness measure tightly related to the bit error probability of the system for an additive white Gaussian noise (AWGN) channel is introduced and the corresponding optimal embedding algorithm described.

III. Orthogonal Dirty Paper Coding

In order to introduce the new orthogonal dirty paper coding technique, let us consider a side information sequence $\mathbf{c}$. Let the length of this sequence be $n = 2^w$, and let $\mathbf{U}$ be a real $n \times n$ unitary matrix such as $\mathbf{U}^T\mathbf{U} = \mathbf{I}_n$. The codebook $\mathcal{U}$ is formed by the n columns of $\mathbf{U}$, i.e., each column of $\mathbf{U}$, say, $\mathbf{u}_i, i = 0, \ldots, n-1$, represents one out of n available codewords. A message of k bits is embedded within a side information block of length n. To this aim, each k-bit message is associated with one codeword that will be referred to as the carrier codeword. Note that since the number of available codewords is n, a clear limit exists for k, i.e., $k \leq \log_2(n)$, or, equivalently, $k \leq w$. Of course, the use of orthogonal codes, somewhat resembling orthogonal modulation and Spread Spectrum systems, is far from optimal from a capacity point of view, since when the length of the code tends to infinity, the code rate tends to zero. Anyway, our analysis is focused on error probabilities rather than capacity, and it mainly applies to practical scenarios that are very far from the asymptotical analysis typical of Inf. Theory. For an in-depth discussion about the differences between performance analysis based on error probability and capacity, see [11] and [21].

As in any QIM scheme, we now must split the codebook $\mathcal{U}$ into a number of subsets (bins), where each is associated with a given message. Given the symmetric structure of our codebook—all the distances between codewords are equal—we can choose any partition of $\mathcal{U}$. Let us then consider 2^k disjoint subsets $Q_l, l = 0, \ldots, 2^k - 1$, such that $\bigcup_l Q_l = \mathcal{U}$, and assume that a one-to-one predefined mapping exists between each possible k-bit information sequence $\mathbf{b}_l, l = 0, \ldots, 2^k - 1$ and the subsets Q_l. This means that each k-bit information sequence can be associated with one out of 2^{w-k} carrier codewords $\mathbf{u}_i$ belonging to the same subset. In the following, we will assume that the lth message has to be hidden within $\mathbf{c}$ and indicate the sequence associated with it by $\mathbf{u}_m$. In order to specify the embedding rule, we must first define the decoding process. In particular, upon receiving $\mathbf{r}$, the decoder estimates the hidden carrier sequence by evaluating

$$\hat{i} = \arg\max_{i=0,\ldots,n-1}(\mathbf{r}^T\mathbf{u}_i) \tag{4}$$

where T stands for transpose operation. The estimated message $\mathbf{b}_{\hat{l}}$ corresponds to the message associated to the bin to which $\mathbf{u}_{\hat{i}}$ belongs. The watermarking signal $\mathbf{w}$ is chosen in such a way that the watermarked feature vector

$$\mathbf{c}_w = \mathbf{c} + \mathbf{w} \tag{5}$$

lies inside the decoding region associated with $\mathbf{u}_m$. In this process, the embedder takes care to reach a given level of robustness with a minimum distortion (see Section III-A). Note that the decoding rule outlined above, together with the equi-energetic nature of the carrier codewords, ensures that the watermark retrieval is invariant with respect to the multiplication by an unknown scale factor.

In Section III-A, we derive an optimal embedding strategy permitting to minimize the embedding distortion for a fixed robustness. To do so, we need to define a proper robustness measure. A possible solution is to directly use the error probability in

the presence of an additive white Gaussian attack, i.e., the error probability resulting by letting $\mathbf{r} = \mathbf{c_w} + \mathbf{n}$ and by assuming that $\mathbf{n} \sim N(0, \sigma_n)$. Minimization of the embedding distortion based on the exact error probability, however, is too complex; hence, we prefer to use the the maximum pairwise error probability between $\mathbf{u}_m$ and the other codewords in $\mathcal{U}$ [22]. The goodness of such a probability as an approximation of the exact error probability is verified experimentally in Section III–B. With the above observations in mind, let us indicate the pairwise error probability between $\mathbf{u}_m$ and $\mathbf{u}_q$ by $P_e(m,q)$; by remembering the decoding rule (4), we have

$$P_e(m,q) = \text{Prob}\{\mathbf{c}_w{}^T(\mathbf{u}_m - \mathbf{u}_q) + z < 0\} \tag{6}$$

where $z = \mathbf{n}^T(\mathbf{u}_m - \mathbf{u}_q) \sim N(0, \sigma_n\sqrt{||\mathbf{u}_m - \mathbf{u}_q||})$, and where the error probability is averaged over $\mathbf{n}$, i.e., by conditioning on the realization of the message $\mathbf{b}_l$ and the cover signal $\mathbf{c}$ (and, hence, $\mathbf{c_w}$). Note that in the sequel (see the next section), we will use (6) to drive the watermark embedding phase; hence, conditioning to $\mathbf{c}$ is a reasonable (necessary) operation, since the embedder wants to fix the robustness of the watermark for any particular realization of $\mathbf{c}$. At the same time, since the embedder will ensure that the same target $P_e(m,q)$ is obtained for every realization of $\mathbf{c}$, $P_e(m,q)$ will also give a measure of the overall robustness of the watermark when averaged over $\mathbf{c}$. The probability (6) can be evaluated by using the approximation[3] [22]

$$P_e(m,q) \cong \frac{1}{2}\exp\left\{-\left[\frac{\mathbf{c}_w{}^T(\mathbf{u}_m - \mathbf{u}_q)}{\sqrt{2}\sigma_n\sqrt{||\mathbf{u}_m - \mathbf{u}_q||}}\right]^2\right\} \tag{7}$$

which can be equivalently expressed as

$$\frac{\mathbf{c}_w{}^T(\mathbf{u}_m - \mathbf{u}_q)}{||\mathbf{u}_m - \mathbf{u}_q||} \cong \sigma_n\sqrt{2 \times \log\left(\frac{1}{2P_e(m,q)}\right)}. \tag{8}$$

Equation (8) can be conveniently rewritten by defining the document-to-noise ratio (DNR) introduced by the attack, i.e.,

$$\text{DNR} = 10\log_{10}\left(\frac{P_c}{\sigma_n^2}\right) \tag{9}$$

where $P_c = E[||\mathbf{c}||^2]/n$ is the power of the side information $\mathbf{c}$. By considering (9) and (8), we obtain

$$\mathbf{c}_w{}^T(\mathbf{u}_m - \mathbf{u}_q) \cong ||\mathbf{u}_m - \mathbf{u}_q|| \times \sqrt{P_c\left(10^{-\frac{\text{DNR}}{10}}\right)} \times \sqrt{2 \times \log\left(\frac{1}{2P_e(m,q)}\right)}. \tag{10}$$

In addition, since $\mathbf{U}^T\mathbf{U} = \mathbf{I}_n$, then $||\mathbf{u}_m - \mathbf{u}_q|| = \sqrt{2}$. Hence, (10) can be expressed as

$$\mathbf{c}_w{}^T(\mathbf{u}_m - \mathbf{u}_q) \cong 2\sqrt{P_c \times \left(10^{-\frac{\text{DNR}}{10}}\right) \times \log\left(\frac{1}{2P_e(m,q)}\right)}. \tag{11}$$

[3]This approximation is known to be good on a log scale.

In the following, we will first derive the optimal embedding rule under a constant robustness constraint (see Section III-A), and then, we will validate the theoretical analysis, and the underlying assumptions, by means of Monte Carlo simulations (see Section III-B)

A. Constant Robustness Embedding

As we outlined in Section II, the informed watermarking paradigm consists of two main steps: choice of the carrier codeword $\mathbf{u}_m$ and embedding of $\mathbf{u}_m$ within $\mathbf{c}$. We first analyze the latter problem and postpone the discussion about the choice of $\mathbf{u}_m$ to the end of this section. Hence, we assume that the carrier codeword $\mathbf{u}_m$ is known and consider the problem of optimally embedding it within $\mathbf{c}$. In particular, we consider a constant robustness optimality criterion for which the watermark robustness is fixed and the embedding distortion Δ, which, in our case, is equal to the watermarking signal energy $||\mathbf{w}||^2$, is minimized. As a measure of robustness, we choose the maximum pairwise message error probability in the presence of Gaussian noise derived in the previous section. Note that since we want a low error probability to be obtained in the presence of Gaussian noise, the marked feature vector surely lies inside the correct decoding region. This in turn guarantees that in the absence of attacks, a null probability of error is obtained, thus ensuring complete host interference rejection.

To go on with the derivation of the optimal embedding procedure, let us denote by

$$P_e^* = \max_q\left\{\frac{1}{2}\exp\left\{-\left[\frac{\mathbf{c}_w{}^T(\mathbf{u}_m - \mathbf{u}_q)}{\sqrt{2}\sigma_n\sqrt{||\mathbf{u}_m - \mathbf{u}_q||}}\right]^2\right\}\right\} \tag{12}$$

the target robustness level, expressed in terms of pairwise error probability [see (7)]. In this case, the constant robustness embedding problem can be formulated as follows. Evaluate the watermarked n-dimensional column vector $\mathbf{c}_w$ or, equivalently, the watermarking signal $\mathbf{w}$, that minimizes the distortion $\Delta = (\mathbf{c}_w - \mathbf{c})^T(\mathbf{c}_w - \mathbf{c})$, subject to the linear constraint[4]

$$\mathbf{c}_w{}^T\mathbf{u}_m - \mathbf{c}_w{}^T\mathbf{u}_q \geq S, \quad \forall \mathbf{u}_q \notin Q_l \tag{13}$$

where

$$S = 2\sqrt{P_c \times \left(10^{-\frac{\text{DNR}}{10}}\right) \times \log\left(\frac{1}{2P_e^*}\right)}. \tag{14}$$

Since the columns of the unitary matrix $\mathbf{U}$ form an orthonormal basis for $\mathbb{R}^n$, it is always possible to express the watermarking signal $\mathbf{w}$ as a linear combination of $\{\mathbf{u}_i\}$, i.e.,

$$\mathbf{w} = \mathbf{U}\mathbf{a} \tag{15}$$

where $\mathbf{a} = (a_0, a_1, \ldots, a_{n-1})^T$ is the column vector with the weights of the linear combination. Given the above, we have

[4]Since we do not impose any upper bound on the allowed distortion, the feasibility set of the linear inequalities in (13) is surely nonempty, as it can be seen by letting $\alpha_i = 0, \forall i \neq m$ and α_m sufficiently high.

$\Delta = \|\mathbf{w}\|^2 = \|\mathbf{a}\|^2 = \sum_{h=0}^{n-1} a_h^2$ and $\mathbf{a}^T\mathbf{U}^T\mathbf{u}_i = a_i$. Accordingly, the information embedding problem is equivalent to finding

$$\mathbf{a} = \arg\min_{\{a_h\}} \left(\sum_{h=0}^{n-1} a_h^2 \right)$$
subject to
$$a_m - a_q + \mathbf{c}^T\mathbf{u}_m - \mathbf{c}^T\mathbf{u}_q \geq S, \quad \forall \mathbf{u}_q \notin Q_l \qquad (16)$$

that is equivalent to

$$\mathbf{a} = \arg\min_{a_m,\{a_q\}} \left(a_m^2 + \sum_{q \,|\, \mathbf{u}_q \notin Q_l} a_q^2 \right)$$
subject to
$$a_q \leq a_m - S + \chi_{q,m}, \quad \forall q \,|\, \mathbf{u}_q \notin Q_l \qquad (17)$$

where $\chi_{q,m} = \mathbf{c}^T\mathbf{u}_m - \mathbf{c}^T\mathbf{u}_q$. Note that a_m in (17) represents the weight of the carrier codeword $\mathbf{u}_m$; hence, it is expected that this term can never be negative. This can be proved by contradiction. Suppose that $a_m < 0$: In this case, the constraint in (17) is also satisfied for $a_m = 0$, which allows us to achieve a lower Δ than any $a_m < 0$, thus contradicting the assumption.

In order to simplify the embedding problem, let us observe that if $a_m - S + \chi_{q,m}$ is greater than or equal to zero, the value of a_q that minimizes Δ while fulfilling the constraint is $a_q = 0$. Conversely, if $a_m - S + \chi_{q,m}$ is lower than zero, then the minimum is obtained on the edge, i.e., for $a_q = a_m - S + \chi_{q,m}$. Hence, the constraint in (17) can be reformulated as

$$a_q = \min(0, a_m - S + \chi_{q,m}). \qquad (18)$$

Accordingly, the minimization problem can be expressed as

$$\mathbf{a} = \arg\min_{a_m} \left(a_m^2 + \sum_{q \,|\, \mathbf{u}_q \notin Q_l} a_q^2 \right)$$
$$a_q = \min(0, a_m - S + \chi_{q,m}), \quad \forall q \,|\, \mathbf{u}_q \notin Q_l. \qquad (19)$$

Note that the problem is now formulated as a one-dimensional minimization problem in the unknown a_m. Such a minimum can be easily computed by means of a numeric approach (e.g., see [23]).

Having defined the optimal embedding rule, we now go back to the choice of $\mathbf{u}_m$, i.e., to the definition of the informed coding mapping associating $\mathbf{b}_l$ with a codeword in Q_l. By recalling that the decoder takes its decision by maximizing the correlation between $\mathbf{r}$ and all the codewords in $\mathcal{U}$, a heuristic way of choosing $\mathbf{u}_m$ consists of choosing the carrier codeword that maximizes the correlation with $\mathbf{c}$, i.e.,

$$\mathbf{u}_m = \arg\max_{\mathbf{u}_s \in Q_l} \mathbf{c}^T\mathbf{u}_s. \qquad (20)$$

Note that this exactly corresponds to the encoding algorithm used in [7] to prove the achievability part of the watermarking channel coding theorem.

B. Simulation Results

The embedding procedure derived in the previous section relies on a couple of simplifying assumptions, namely, the use of the pairwise error probability and the approximation (7); hence, its validity has been checked through Monte Carlo simulations. During the simulations, both the side information $\mathbf{c}$ and the external attack have been generated according to an AWGN model. First of all, we verified the validity of the analysis leading to (7)–(11). In particular, we verified whether the actual error rate corresponds to that predicted theoretically on the basis of the imposed robustness level. Table I shows the actual message error probability P_e obtained through simulations for DNR $=$ 10 dB and for different target robustness levels, namely, $P_e^* = 10^{-2}, 10^{-3}$, and 10^{-4} [remember that, according to (14), we chose to measure robustness in terms of the maximum pairwise error probability, since we argued that this is a good approximation of the true error probability P_e]. In particular, the message error rate P_e obtained through simulations represents the probability that the estimated sequence $\mathbf{b}_{\hat{l}}$ is different from the actual information sequence embedded in $\mathbf{c}_w$. The actual error probabilities have been obtained by averaging over $\mathbf{c}$ the noise sequence and the hidden message. It is readily seen that for both $k = 1$ and $k = 2$, the actual error probability P_e is slightly lower than the target values, thus confirming the validity of the proposed analytical approach. Note also that the lower the probability, the closer P_e is to P_e^*. This is due to the approximation in (7), which becomes more accurate for higher signal-to-noise ratios, i.e., for lower error probabilities.

TABLE I
MESSAGE ERROR PROBABILITY P_e OBTAINED THROUGH SIMULATIONS FOR DNR $=$ 10 dB, FOR DIFFERENT VALUES OF P_e^* [SEE (14)] AND FOR $k = 1$ AND $k = 2, \mathrm{n} = 32$

	$k = 1$	$k = 2$
$P_e^* = 0.01$	0.0063	0.0072
$P_e^* = 0.001$	0.00071	0.0008
$P_e^* = 0.0001$	0.000087	0.000093

Then, we evaluated the overall performance of the proposed approach as a function of the watermark-to-noise ratio (WNR), which is defined as

$$\mathrm{WNR} = \frac{E[\|\mathbf{w}\|^2]}{E[\|\mathbf{r} - \mathbf{c}_\mathbf{w}\|^2]}. \qquad (21)$$

The plots have obtained by fixing the watermark strength, measured in terms of document-to-watermark ratio (DWR), i.e.,

$$\mathrm{DWR} = \frac{E[\|\mathbf{c}\|^2]}{E[\|\mathbf{w}\|^2]}. \qquad (22)$$

A summary of the results we obtained is given in Fig. 4, where the bit error probability is plotted as a function of WNR. The results refer to $n = 32$, different values of DWR (14 and 18 dB), and $k = 1$ and $k = 2$ (i.e., $R = 1/16$ and $R = 1/32$).

To better appreciate the effectiveness of the orthogonal dirty paper coding scheme, we compared it against the ISS algorithm described in [18]. As briefly outlined in the introduction, this scheme is somewhat similar to the one presented here, since it is still based on the informed embeddign paradigm, and it is invariant against the gain attack. The main difference between

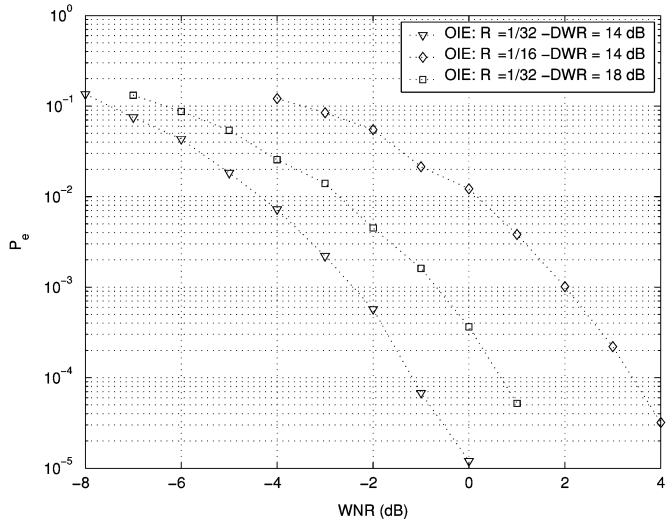


Fig. 4. Orthogonal Informed Embedding (OIE) performance: $n = 32, k = 1 (R = 1/32)$, and $k = 2 (R = 1/16)$.

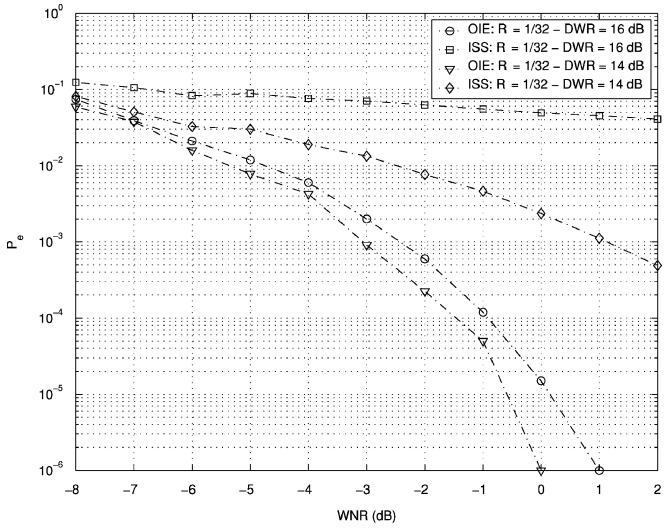
the two schemes is that ISS does not make a provision for informed coding, since the codeword associated with each message is fixed. As it can be seen in Fig. 5, this represents a great penalty for ISS, which, for the payloads and DWRs reported in the figures, always performs worse than our scheme. As a matter of fact, ISS performs better than our algorithm only for very low bit rates $(R < 1/100)$ and/or low DWR.

IV. Quasi-Orthogonal Dirty Paper Coding

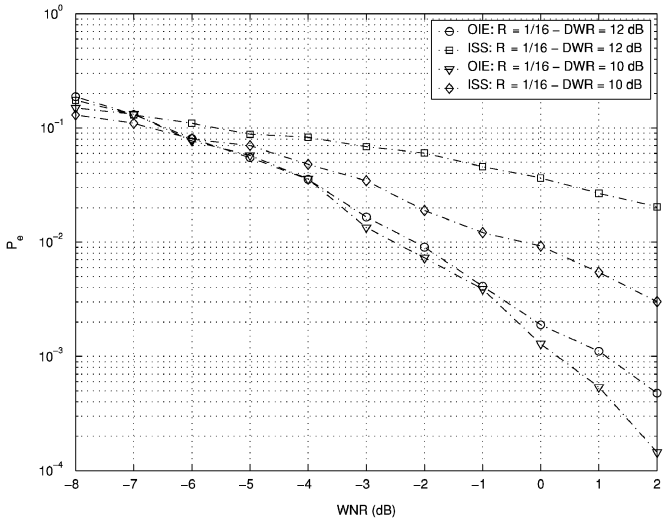
To improve the performance of the proposed system, we start by replacing the orthogonal codes with quasi-orthogonal sequences to increase the number of available codewords for a given sequence length n. Specifically, we will use Gold sequences of length n since their cross-correlation properties ensure that different sequences are almost orthogonal among them [22], [24]. We start by reviewing the embedding procedure derived so far, and then, we will evaluate the performance improvement by means of numerical simulations (see Section IV-B).

A. Constant Robustness Embedding

In this section, we will assume that the column vectors $\mathbf{u}_i$ of the matrix $\mathbf{U}$ are Gold sequences [22], [24]. In this case, the matrix $\mathbf{U}$ is a rectangular $n \times h$ matrix with column vectors $\mathbf{u}_i, i = 1, \ldots, h$, representing a set of h Gold sequences with length n. Gold sequences have been widely studied in the technical literature, particularly for spread spectrum applications, since their autocorrelation and cross-correlation functions are reminiscent of the properties of white noise. Specifically, in the following, we will assume that $\mathbf{u}_i$ are normalized Gold sequences with $u_i(l) = \pm(1/\sqrt{n}), \forall i, l$. Note that all Gold sequences have the same norm, thus ensuring that the decoder performance is invariant with respect to multiplication by a gain factor g. In this case, for a given length $n = 2^w - 1$, the number of possible Gold sequences that are characterized by good periodic cross-correlation properties is $n+2$. Since each cyclic shift



(a)



(b)

Fig. 5. OIE versus ISS performance. (a) $n = 32, k = 1$ (OIE), DWR $= 14$, and 16 dB. (b) $n = 16, k = 1$ (OIE), DWR $= 10$ dB, and 12 dB.

of any Gold sequence is still characterized by the same properties, the overall number of Gold sequences that can be considered for information embedding is $h = n(n + 2)$. Note that, as required to write (15), Gold sequences constitute a frame for $\mathbb{R}^n$, hence ensuring that every element of $\mathbb{R}^n$ can be expressed as a linear combination of the $\mathbf{u}_i$'s.

Let us now consider the distortion introduced by watermark embedding. We have

$$\Delta = \left\| \sum_{i=1}^{h} a_i \mathbf{u}_i \right\|^2 = \sum_{i=1}^{h} a_i^2 + \sum_{i \neq j} a_i a_j \mathbf{u}_i^T \mathbf{u}_j. \qquad (23)$$

We now argue that, due to the particular properties of Gold sequences, the first term of the above equation is dominant with respect to the second, even if the second term is not exactly equal to zero due to the nonperfect orthogonality of $\mathbf{u}_i$'s. Such an assumption has been validated through numerical simulations, whereby we verified that at least for the values of n and DWR used throughout this paper, on the average, the ratio between the second and the first term in the right part of (23) is about -7 dB.

TABLE II
GIE ACTUAL BLOCK ERROR PROBABILITY P_e OBTAINED THROUGH SIMULATIONS FOR DNR $= 10$ dB, FOR DIFFERENT VALUES OF P_e^* [SEE (14)] AND FOR $k = 1$ AND $k = 2, \mathrm{n} = 32$

	$k=1$	$k=2$
$P_e^* = 0.01$	0.02	0.023
$P_e^* = 0.001$	0.004	0.0043
$P_e^* = 0.0001$	0.00083	0.00091

Moreover, with very high probability, the second term is lower than zero, hence contributing to the increase of the DWR. By relying on the above observations, the minimization of $\sum_{i=1}^{m} a_i^2$ can still be considered to be the target of the information embedding problem.

We must now consider the impact of the nonperfect orthogonality of Gold sequences on the robustness constraint. To this aim, let us consider the term

$$D(m,q) = \mathbf{c}_w{}^T\mathbf{u}_m - \mathbf{c}_w{}^T\mathbf{u}_q \tag{24}$$

which gives a measure of the robustness of the information embedding procedure. Since $\mathbf{c}_w = \mathbf{c} + \mathbf{w}$ and $\mathbf{w} = \sum_{i=1}^{m} a_i\mathbf{u}_i$, (24) can be expressed as

$$D(m,q) = \mathbf{c}^T\mathbf{u}_m - \mathbf{c}^T\mathbf{u}_q + a_m\sum_{l=1}^{h} u_m^2(l) - a_q\sum_{l=1}^{h} u_q^2(l) + \sum_{l\neq m} a_l\mathbf{u}_l^T\mathbf{u}_m - \sum_{l\neq q} a_l\mathbf{u}_l^T\mathbf{u}_q. \tag{25}$$

Since $u_i(l) = \pm(1/\sqrt{n})$, we get directly from (25)

$$D(m,q) = \mathbf{c}^T\mathbf{u}_m - \mathbf{c}^T\mathbf{u}_q + a_m - a_q + \sum_{l\neq m} a_l\mathbf{u}_l^T\mathbf{u}_m - \sum_{l\neq q} a_l\mathbf{u}_l^T\mathbf{u}_q. \tag{26}$$

By invoking again the cross-correlation properties of Gold sequences, we argue that the last two terms in (26) have a minor impact on the bit error probability expression. Of course, this is not completely true; hence, it is expected that for a given threshold S [computed by neglecting the two additional terms in (26)], the error probability P_e will slightly increase with respect to the orthogonal case.

B. Simulation Results

As we noted in the previous section, the adoption of Gold sequences instead of truly orthogonal codes results in the appearance of an additional noise term, that, according to (26), depends on the unknown weights a_l. Hence, the actual error probability P_e cannot be given *a priori* but can only be estimated by means of computer simulations.

As an example, Table II shows the actual error probability P_e obtained through simulations for the Gold Informed Embedding (GIE) strategy described above. Results in Table II are derived in the case of DNR $= 10$ dB and for different target robustness levels, i.e., $P_e^* = 10^{-2}, 10^{-3}, 10^{-4}$. Note that, as expected, in this case, the actual error probabilities are higher than P_e^*, on account of the additional noise term that is not included in the evaluation of S in (14). However, for a given (actual) P_e, due to the higher number of available sequences in $\mathcal{U}$, the pseudo-noise informed embedding strategy permits us to noticeably increase the DWR, hence resulting in better overall performance.

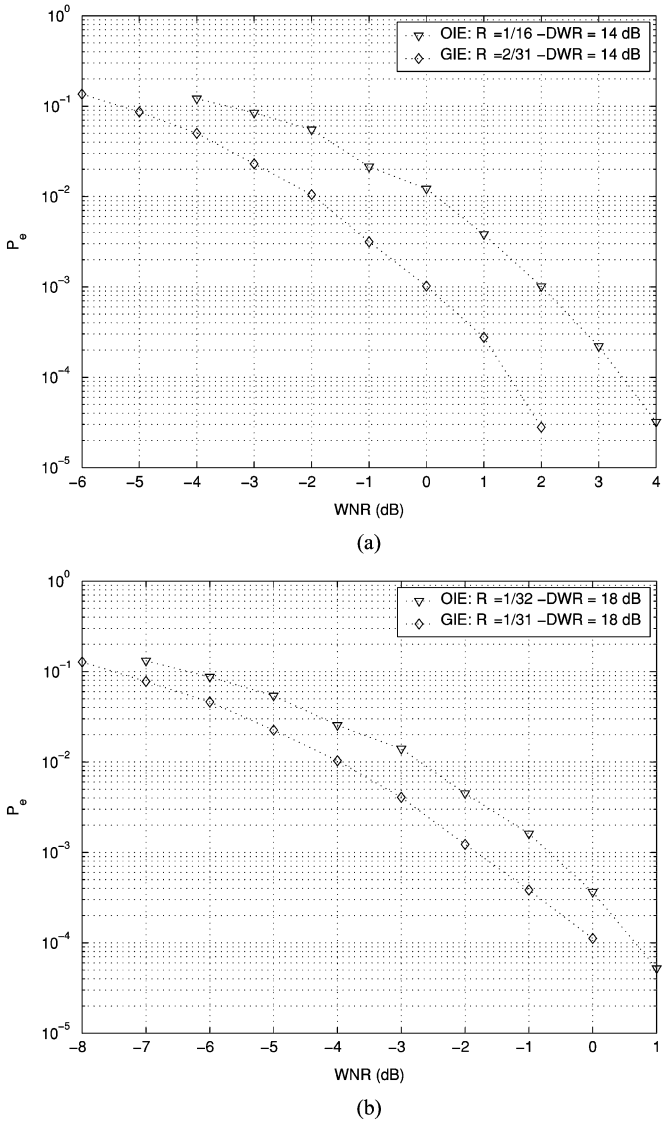


Fig. 6. OIE versus GIE performance. (a) $n = 32$ (OIE), $n = 31$ (GIE), $k = 1$ and $k = 2 (R = 1/16, R = 2/31)$, DWR $= 14$ dB. (b) DWR $= 18$ dB.

Such performances, resulting from Monte Carlo simulations, are shown in Fig. 6 and for $R = 2/31 (n = 31, k = 2)$ and DWR $= 14$ dB in part (a) of the figure and $R = 1/31 (n = 31, k = 1)$ and DWR $= 18$ dB in part (b). For comparison purposes, the Orthogonal Informed Embedding (OIE) curves for $R = 1/16$ [$n = 32, k = 2$, part (a)] and $R = 1/32$ [$n = 32, k = 1$, part (b)] are also given. Note that the GIE strategy allows us to get an improvement ranging from 1 to 2 dB. It is worth noting that, on account of Gold sequence construction constraints, in the Gold case, we have $n = 2^w - 1$, i.e., it is not possible to have exactly the same length for the GIE and OIE cases. Hence, the actual performance gain of GIE is slightly better than that shown in Fig. 6.

V. Multistage Decoding

As a second way to improve the performance of the basic scheme described in Section III, we insert an additional channel coding step prior to orthogonal (or Gold) dirty paper coding.

To start with, let us remember that the proposed informed embedding strategy takes a block of k bits and associates them with the transmitted sequence $\mathbf{c_w}$ of length n in a similar way to a classical one-step channel coding scheme. Hence, following the analogy with communication systems, the overall system performance can be improved by means of a serial concatenation of channel codes. In particular, we consider an outer code with rate $R_c = k_0/k$ that takes a block of N_b bits at its input (i.e., the watermark message) and gives a block of N_b/R_c bits at its output. Such bits are then taken k at a time by the informed embedding scheme (that acts as the inner code), thus producing $(N_b/k_0)\times n$ samples at its output. Hence, in the proposed scheme, the side information $\mathbf{c}$ is given by $(N_b/k_0) \times n$ samples, providing an overall rate of k_0/n.

The detection strategy (4) generates hard estimates of the bits $\mathbf{b}_l = (b_{l,0}, b_{l,1}, \ldots, b_{l,k-1})$. On the other hand, when dealing with multistage decoding, it is preferable that the inner decoder produce soft estimates to be delivered to the outer decoder [22]. To do so, let us introduce the sets $I_{1,s}$ and $I_{0,s}$ as

$$I_{1,s} = \{l : b_{l,s} = 1\}$$
$$I_{0,s} = \{l : b_{l,s} = 0\}. \tag{27}$$

More specifically, $I_{1,s}(I_{0,s})$ represents the set of 2^{k-1} sequences $\mathbf{b}_l$ for which the sth bit is 1 (0). Then, by focusing on the sth bit ($s = 0, \ldots, k-1$), we consider the following reliability measure of the estimated value of $b_{l,s}$:

$$v_s = \max_{\mathbf{u}_i \in Q_l, l \in I_{1,s}} (\mathbf{r}^T \mathbf{u}_i) - \max_{\mathbf{u}_i \in Q_l, l \in I_{0,s}} (\mathbf{r}^T \mathbf{u}_i). \tag{28}$$

In practice, among all the carrier codewords associated with a message with a 1 in the sth position, we pick up the sequence having the maximum correlation with the received vector $\mathbf{r}$. Then, we repeat the same procedure for the codewords corresponding to the messages with a 0 in the sth position and take the difference between the two (maximum) correlation values obtained in this way as a soft estimate of the sth bit of $\mathbf{b}_l$. In summary, the sign of (28) determines the hard estimate of the sth bit, and its absolute value represents the soft output information that can be used by the outer decoder.

It is worth pointing out that the soft first-stage detection strategy presented in the previous paragraphs can be applied to any kind of binary outer coder's structure. In this paper, the outer code is the $R_c = 1/2$ binary punctured parallel concatenated turbo coder presented in [25]. Note that, given the structure of the outer code, we have $k_0 = 1$ and $k = 2$, i.e., the informed embedding strategy works by taking blocks of 2 bits at a time. Moreover, the overall rate of the multistage scheme is equal to $1/n$.

A. Simulation Results

We first evaluate the improvement brought by the insertion of a turbo code on top of the inner dirty coding scheme. Such an improvement is evident if we look at the plot given in Fig. 7, where the performance of Turbo-Coded GIE (TC-GIE) are compared

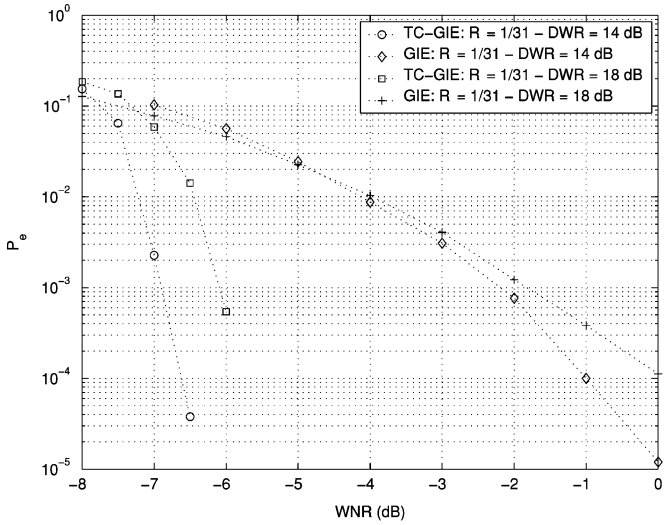


Fig. 7. GIE versus Turbo Coded-GIE (TC-GIE) performance. $n = 31, k = 1$ (GIE). $k = 2$ (TC-GIE).

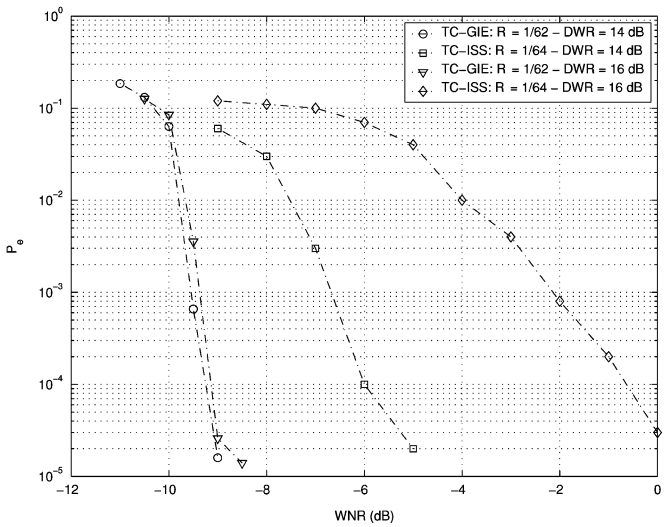
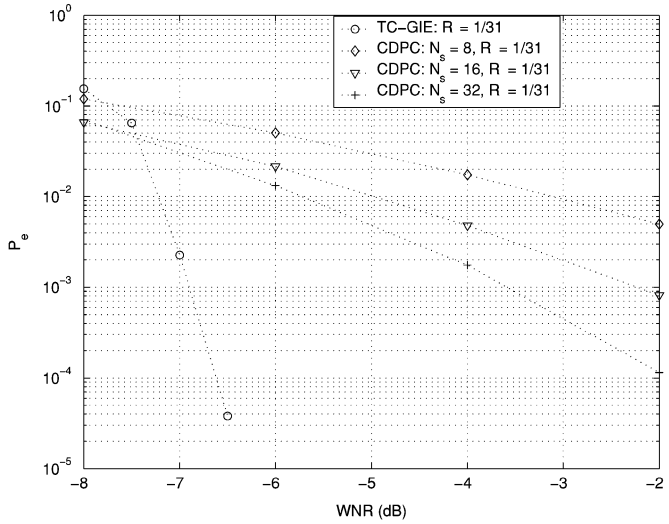


Fig. 8. Comparison between TC-GIE and TC-ISS.

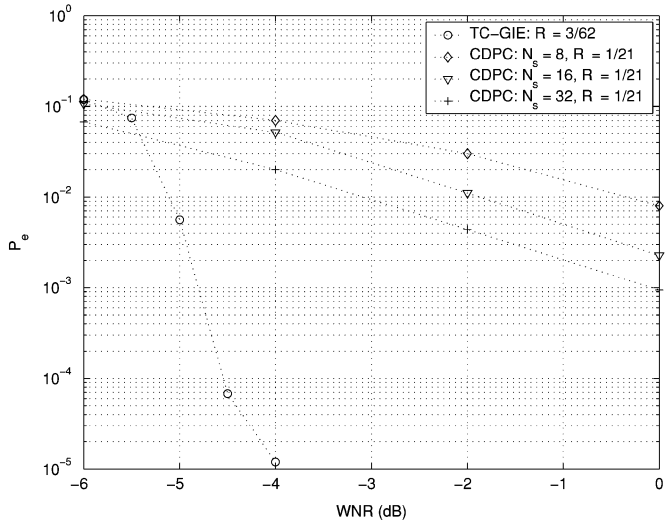
with those of the plain GIE scheme for two different DWRs, i.e., DWR $= 14$ and DWR $= 18$ dB. In both cases, the overall rate was set to $R = 1/31$; however, in the TC-GIE case, the rate of inner dirty paper code is $2/31(n = 31, k = 2)$ with an additional $1/2$ factor due to the outer turbo code. In contrast, for the GIE case, a dirty paper code with $R = 1/31(n = 31, k = 1)$ was used.

In order to get a more precise idea of the overall value of TC-GIE, we compared it against a turbo-coded version of ISS (TC-ISS) and the dirty trellis scheme proposed by Miller *et al.* [14], [15]. Fig. 8 gives the results we obtained when comparing TC-GIE against TC-ISS. As it can be seen, the superior potentialities of our approach, already highlighted in Fig. 5, are confirmed even when channel coding is taken into account.

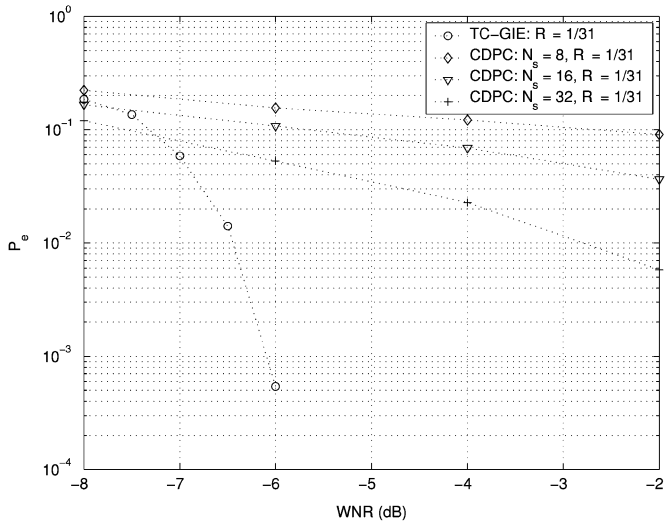
The second comparison we made regarded TC-GIE and Miller's system [14], [15]. In Figs. 9(a)–(c), the bit error probability for the TC-GIE scheme and the Convolutional Dirty Paper Coding (CDPC) approach described in [14] and [15] is shown, for different values of the overall bit rate. In Fig. 9(a)



(a)



(b)



(c)

Fig. 9. TC-GIE versus CDPC performance. (a) and (b) DWR $=$ 14 dB. (c) DWR $=$ 18 dB.

and (b), the DWR is set to 14 dB, whereas in Fig. 9(c), we have considered DWR $=$ 18 dB. Results refer to an AWGN channel

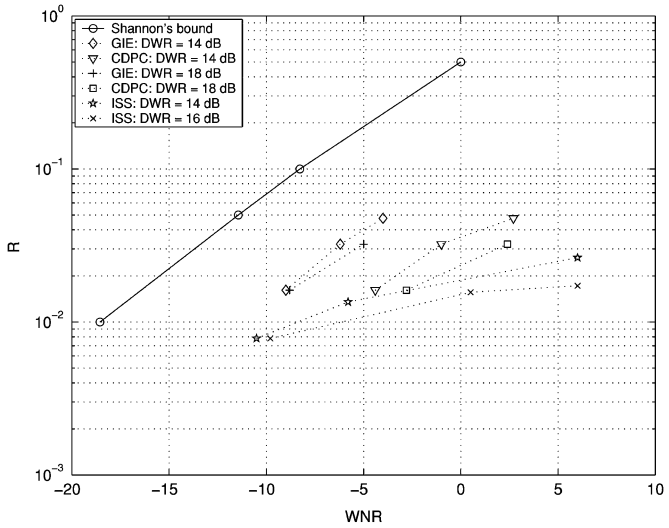


Fig. 10. TC-GIE, TC-ISS, and CDPC capacity curves.

with no gain attack since it is known in advance that both schemes are intrinsically robust against value-metric scaling. The CDPC curves have been obtained for different numbers of transitions per state,[5] which we let to be equal to the number of states N_s. As to the number of states, we have considered the cases $N_s = 8, N_s = 16$, and $N_s = 32$, yielding coding structures with increasing computational burden. Indeed, in the CDPC scheme, the transmitted signal $\mathbf{c}_w$ is evaluated by means of an iterative approach [15]. At each iteration, a Viterbi decoder is run whose complexity depends on the number of states and transitions of the convolutional code. An exact evaluation of the complexity of the CDPC scheme is a cumbersome task, since there is no way to evaluate in advance the number of steps requested for convergence. However, a fair comparison among CDPC, TC-OIE, and TC-GIE, in terms of computational effort, can be given by considering the computing time requested for watermark embedding. In particular, for a watermark length $N_b = 1000$, such a time was of a few seconds in the TC-OIE and TC-GIE cases, whereas in the CDPC case, it went from nearly 20 min (for $N_s = 8$) up to several hours (for $N_s = 32$). Hence, it is possible to state that the proposed informed embedding strategy allows us to noticeably reduce the implementation complexity with respect to the dirty trellis approach.

As a last result, we considered the performance of TC-GIE from a capacity point of view in that we compared the bit rates achieved by TC-GIE against the capacity limit predicted by theory. To do so, we set the bit error probability to 10^{-5} and measured the corresponding values of WNR and R. The results we obtained for TC-GIE, TC-ISS, and CDPC schemes are depicted in Fig. 10. The number of states is set to $N_s = 32$ for the CDPC case. Once again, the superiority of the new scheme comes out, even if the capacity limit is somewhat far away. This is not surprising since Gold codes are known to have a poor error-correcting capability, which is a drawback that is only partially mitigated by the presence of the outer turbo code. The general trend of TC-ISS, whose performance tends to be improve for very low bit rates, can also be appreciated.

[5]In the CDPC scheme, the number of transitions per state determines the code's dirtiness.

VI. Conclusion

By relying on the simple structure of orthogonal and Gold sequences, we have presented a new dirty paper coding watermarking scheme. A first advantage of the new scheme with respect to other informed watermarking algorithms such as QIM and ST-DM is that due to the equi-energetic nature of the codewords and to the adoption of a correlation-based decoder, robustness against value-metric scaling is automatically achieved. A second merit of the proposed scheme is the use of an optimal embedding strategy, which permits a decrease in the distortion introduced by the watermark for a fixed message error rate. Finally, the new scheme is very simple, thus allowing very fast watermark embedding. We have also shown how the performance of the system is dramatically improved by concatenating the dirty paper code with an outer turbo code. To this aim, we had to introduce a new soft dirty paper decoding scheme that allows the iterative multistage decoding of the concatenated codes. The validity of the proposed techniques have been assessed through Monte Carlo simulations.

Several directions for future work remain open. First of all, we are planning to apply orthogonal and pseudo-random dirty paper coding to the watermarking of real data, namely, still images and video. We are also working on a modified version of the embedding algorithm, whereby the embedding problem is reformulated in such a way that the bit error probability is minimized for a given distortion. This way of operating would be particularly useful to design a perceptually based dirty paper coding scheme, where the maximum allowable distortion for each host sample is determined through perceptual considerations. Finally, the possibility of using more powerful spherical codes [13], [16], [17] than the simple orthogonal codes used in this paper is under investigation.

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

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

[4] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," in *Proc. IEEE Int. Symp. Inf. Theory*, Sorrento, Italy, Jun. 2000, p. 19.

[5] P. Moulin, "The role of information theory in watermarking and its application to image watermarking," *Signal Process.*, vol. 81, no. 6, pp. 1121–1139, 2001.

[6] B. Chen and G. Wornell, "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.

[7] A. S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1639–1667, Jun. 2002.

[8] W. Yu, A. Sutivong, D. Julian, T. M. Cover, and M. Chiang, "Writing on colored paper," in *Proc. ISIT*, Washington, DC, Jun. 24–29, 2001.

[9] A. S. Cohen and A. Lapidoth, "Generalized writing on dirty paper," in *Proc. ISIT*, Lausanne, Switzerland, Jun. 30–Jul. 5 2002.

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

[11] F. Perez-Gonzalez, F. Balado, and J. R. Hernandez, "Performance analysis of existing and new methods for data hiding with known-host information in additive channels," *IEEE Trans. Signal Processing*, vol. 51, no. 4, pp. 960–980, Apr. 2003.

[12] J. J. Eggers, R. Bäuml, and B. Girod, "Estimation of amplitude modifications before SCS watermark detection," *Proc. SPIE Security Watermarking Multimedia Contents IV*, vol. 4675, pp. 387–398, Jan. 2002.

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

[14] I. J. Cox, M. L. Miller, and G. J. Doerr, "Dirty-paper trellis codes for watermarking," in *Proc. 9th IEEE Int. Conf. Image Process.*, vol. II, Rochester, NY, Sep. 2002, pp. 129–132.

[15] 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.

[16] J. Hamkins and K. Zeger, "Asymptotically dense spherical codes—Part I: Wrapped spherical codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1774–1785, Nov. 1997.

[17] ——, "Asymptotically dense spherical codes—Part II: Laminated spherical codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1786–1798, Nov. 1997.

[18] H. S. Malvar and D. A. F. Florencio, "Improved spread spectrum: A new modulation technique for robust watermarking," *IEEE Trans. Signal Processing*, vol. 51, no. 4, pp. 898–905, Apr. 2003.

[19] A. El Gamal and T. M. Cover, "Multiple user Inf. Theory," *Proc. IEEE*, vol. 68, no. 12, pp. 1466–1485, Dec. 1980.

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[21] F. Balado-Pumarino, "Digital Image Data Hiding Using Side Information," Ph.D. dissertation, Univ. Vigo, Vigo, Spain, 2003.

[22] J. G. Proakis, *Digital Communications*, Second ed. New York: McGraw-Hill, 1989.

[23] G. E. Forsythe, M. A. Malcolm, and C. B. Moler, *Computer Methods for Mathematical Computations*. New York: Prentice-Hall, 1976.

[24] R. Gold, "Optimal binary sequences for spread spectrum multiplexing," *IEEE Trans. Inf. Theory*, vol. IT-13, pp. 619–621, Oct. 1967.

[25] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correcting coding and decoding: Turbo-codes," in *Proc. IEEE Int. Conf. Commun.*, Geneva, Switzerland, May 1993, pp. 1064–1070.

Andrea Abrardo (M'99) received the Dr.Ing. degree in electronic engineering from the University of Florence, Florence, Italy, in April 1993. In June 1998, he received the Ph.D. degree from the University of Florence, where his dissertation focussed 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 networks infrastructures. Since August 1998, he has been with the Department of Information Engineering, the 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.

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 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 Signal Processing Letters and the IEEE Signal Processing Magazine (Column and Forum section). Dr. Barni is a member of the IEEE Multimedia Signal Processing Technical Committee (MMSP-TC).