

ON THE OPTIMAL STRUCTURE OF WATERMARK DECODERS UNDER DESYNCHRONIZATION ATTACKS

Pierre Moulin

University of Illinois
Beckman Inst., Coord. Sci. Lab & ECE Dept.
405 N. Mathews Ave., Urbana, IL 61801
Email: moulin@ifp.uiuc.edu

ABSTRACT

Designing watermarking codes that can withstand geometric and other desynchronization attacks is a notoriously difficult problem. One may ask whether these difficulties are due to limitations of current codes, or rather to fundamental limitations on achievable performance. This paper describes our recent results on this problem for blind and nonblind watermarking, and provides examples for which the theory applies.

Keywords: watermarking, data hiding, desynchronization, decoding, information theory, image processing.

1. INTRODUCTION

One of the main difficulties in designing watermarking and data-hiding codes is to ensure a certain level of robustness against desynchronization and other geometric attacks. Such attacks include image warping, amplitude modulation, and for audio and video signals, temporal desynchronization [1]–[4]. If the original host signal is available to the decoder (*nonblind watermarking*), there is clearly hope to “undo” these attacks with the help of this signal, and the problem may be broadly viewed as one of image registration. In the opposite situation where the host is not available to the decoder (*blind watermarking*), the task seems much harder. One may wonder whether this increased difficulty is due to the deficiencies of current code designs, or to some fundamental performance limit. One may hope that analogously to the Gaussian watermarking and dirty-paper coding problems [5], there is no loss of performance relative to the nonblind case.

In this paper, we describe our recent results in that direction, starting from a mathematically tractable formulation of the problem. The host signal is modeled as Gaussian. The attack is modeled as the cascade of a Gaussian channel and a smooth, invertible mapping representing the geometric attack. This mapping is parameterized by an unknown parameter θ (e.g., a scaling parameter, a filter response, or a time-warping function). We address the potential loss in error probability due to lack of knowledge of θ by the receiver.

The framework for this study is *universal decoding* [6]. Roughly speaking, a decoder is said to be universal if it performs as well (in terms of error exponents) as a coherent decoder that *knows* θ . For many problems, universal decoders do not exist. In other problems (roughly speaking, when the complexity of the family of channel distributions parameterized by θ is sufficiently small), universal decoders may exist. For our problem, this notion of complexity can be

reduced to a notion of complexity for the family of mappings. In this sense, delay attacks, scaling attacks, and even some warping attacks, turn out to be “simple”.

In this paper we introduce this framework with its underlying assumptions, and state our main results. (Proofs are given in [7].) While the structure of a universal decoder may be prohibitively complex [6], under our attack model we obtain two interesting results:

1. For nonblind watermarking, the encoder uses a spread-spectrum scheme. The universal decoder takes the form of a *joint estimator/decoder* in which the cost function is the *normalized correlation coefficient* between the received data and codebook elements.
2. For blind watermarking, the encoder is a lattice Quantization Index Modulation (QIM) scheme [8, 9]. The universal decoder takes the form of a *joint estimator/decoder* in which the decoder is a minimum-distance decoder.
3. The same random codebooks that are effective in the absence of desynchronization attacks remain ideal.

Regarding Point #1, the use of normalized correlation as a decoding metric is by no means new (it has long been known to be an effective choice against scaling attacks), and so one contribution of this paper is to establish the optimality properties of this decoding metric under a much broader setting.

Regarding Point #3, the claim applies to random codebooks and not to any specific family of structured codebook.

Notation. We use uppercase letters to denote random variables, lowercase letters to denote their individual values, and boldface letters to denote sequences of real numbers, e.g., $\mathbf{x} = (x_1, \dots, x_n)$. Also, $\|\mathbf{x}\| = (\sum_i x_i^2)^{1/2}$ denotes the Euclidean norm of \mathbf{x} . The Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$. The symbol $f(n) \doteq g(n)$ denotes asymptotic equality on the logarithmic scale, i.e., $\lim_{n \rightarrow \infty} \frac{1}{n} \ln f(n)/g(n) = 0$.

2. MATHEMATICAL MODEL

We adopt the following communication model for watermarking, see Fig. 1 [10]. Data are to be embedded at a rate of R bits per sample (pixel) in a host signal (image). Given a host sequence $\mathbf{s} \in \mathbb{R}^n$, side information $\mathbf{k} \in \mathcal{K}^n$, and a message $m \in \{1, \dots, \lceil 2^{nR} \rceil\}$, the encoder produces a marked sequence $\mathbf{x} = f_n(\mathbf{s}, m, \mathbf{k})$ where f_n is the encoding function. The marked sequence \mathbf{x} is subject to attacks, resulting in a degraded sequence \mathbf{y} . The decoder returns an estimate $\hat{m} = \psi_n(\mathbf{y}, \mathbf{k})$ of the message that was sent. The side information \mathbf{k} may be a cryptographic key, independent of \mathbf{S} (*blind watermarking*);

\mathbf{k} may also contain full information about \mathbf{S} (*nonblind watermarking*).

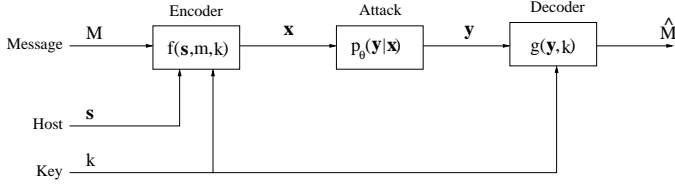


Fig. 1. Communication model for watermarking and data hiding.

Referring to Fig. 2, the attack is modeled by the cascade of a fixed memoryless channel $W(z|x)$ and an invertible global mapping T_θ representing a desynchronization attack. Therefore $\mathbf{y} = T_\theta \mathbf{z}$, where \mathbf{z} is generated according to the product probability density function (pdf) $W^n(\cdot|x)$. There are variations on this problem, in which T_θ is noninvertible, but these variations will not be considered in this paper.

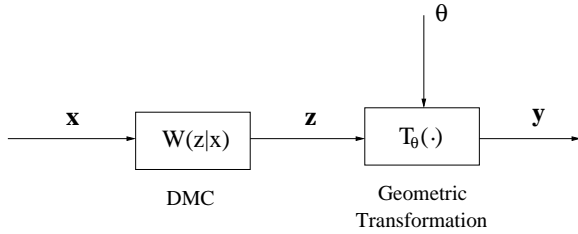


Fig. 2. Model for geometric attacks.

We focus on the Gaussian problem where the host \mathbf{S} is a sequence of independent, identically distributed (iid) Gaussian random variables with marginal pdf $\mathcal{N}(0, \sigma_S^2)$. The channel $W(z|x)$ is Gaussian and may be written in the form

$$Z = X + N \quad (1)$$

where $N \sim \mathcal{N}(0, \sigma_N^2)$ is independent of X , the channel input.

Our assumptions on the family $\{T_\theta, \theta \in \Theta_n\}$ are listed below.

(A1) The mapping $T_\theta : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is invertible for all n and for all $\theta \in \Theta_n$.

We denote by $U_\theta = T_\theta^{-1}$ the inverse mapping. In addition, the parameter set satisfies one of the conditions (A2) or (A2') below, the latter being a relaxed version of the former.

(A2) The parameter set Θ_n is discrete, and its cardinality is fixed or grows subexponentially with n :

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \ln |\Theta_n| = 0.$$

(A2') There exists a sequence $\epsilon_n \downarrow 0$ and a sequence of sets $\tilde{\Theta}_n \subseteq \Theta_n$ that satisfies the following two conditions. First, the cardinality of these subsets is upper bounded by a subexponential function of n :

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \ln |\tilde{\Theta}_n| = 0.$$

Second, the collection of sets $\tilde{\Theta}_n$ is dense in Θ_n in the following sense. For any $\mathbf{z} \in \mathbb{R}^n$ and $\theta \in \Theta_n$, one can find some $\theta^* \in \tilde{\Theta}_n$ such that

$$\|\mathbf{z} - U_{\theta^*} T_\theta \mathbf{z}\| \leq \epsilon_n \|\mathbf{z}\|. \quad (2)$$

3. EXAMPLES

Three examples of mappings T_θ are provided below.

3.1. Block Permutations

Let $\theta : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$ denote a permutation of the samples $\{1, 2, \dots, n\}$, and $T_\theta : \mathbb{R}^n \rightarrow \mathbb{R}^n$ the corresponding permutation operator. The invertibility condition **(A1)** is automatically satisfied. If we choose Θ_n to be the set of *all* permutations, then Θ_n has size $|\Theta_n| = n! \doteq (n/e)^n$ (by Stirling's formula), which is superexponential in n ; conditions **(A2)** and **(A2')** are violated.

To make Θ_n less complex, we could choose Θ_n to be the set of all block permutations, for blocks of size B . The size of Θ_n is now

$$|\Theta_n| = (n/B)! \doteq \left(\frac{n}{eB}\right)^{n/B}$$

which is subexponential in n (i.e., **(A2)** holds) if and only if $\ln |\Theta_n| \ll n$, or equivalently, $B \gg \ln n$.

3.2. Scaling

Consider amplitude scaling with log scaling parameter $\theta \in [0, 1]$. That is, $\mathbf{y} = T_\theta \mathbf{z} = e^\theta \mathbf{z}$. The parameter θ belongs to an interval, say $\Theta_n = [0, 1]$. Choose an integer sequence K_n , subexponential in n , and define $\tilde{\Theta}_n = \{k/K_n, 1 \leq k \leq K_n\}$, a uniform discretization of Θ_n . Therefore, given any $\theta \in \Theta_n$ and $\mathbf{z} \in \mathbb{R}^n$, we can find $\theta^* \in \tilde{\Theta}_n$ such that

$$\|\mathbf{z} - U_{\theta^*} T_\theta \mathbf{z}\| = |1 - e^{\theta - \theta^*}| \|\mathbf{z}\| \leq (1 - e^{-1/K_n}) \|\mathbf{z}\| \leq \frac{1}{K_n} \|\mathbf{z}\|.$$

Hence (2) holds with $\epsilon_n = 1/K_n$.

3.3. Linear Time-Invariant Filtering

Let $\theta = (\theta_1, \dots, \theta_n)$ be a length- n sequence, and T_θ denote the circular convolution operator:

$$\mathbf{y} = T_\theta \mathbf{x} \triangleq \theta \star \mathbf{x}.$$

The invertibility condition **(A1)** is satisfied, for instance, when

- θ is a odd-length filter whose taps are symmetric around $i = 1$ and satisfy the condition $|\theta_1| > 2 \sum_{i=2}^{(n+1)/2} |\theta_i|$.
- θ is a circular shift parameter:

$$(T_\theta x)_i = \sum_{j=1}^n x_j \varphi(i - j - \theta)$$

where $\varphi(t) = \frac{\sin \pi t}{n \sin \pi t/n}$ is the periodic sinc interpolating function. If θ is an integer, the above formula simplifies to $(T_\theta x)_i = x_{i - \theta \bmod n}$.

- T_θ is the cascade of these two operators.

Unlike the permutation examples of Sec. 3.1, here the feasible set Θ_n is a continuum. The complexity of Θ_n can be controlled, for instance, by introducing a d -dimensional parametric representation of the scale-normalized filter, where $d \ll n$, and quantizing the coefficients in this representation, analogously to Sec. 3.2. Then Assumption **(A2')** holds.

4. COMPOSITE HYPOTHESIS TESTING

The decoding problem is a M -ary hypothesis testing problem of the form

$$H_m : \mathbf{Y} \sim p_\theta(\mathbf{y}, \mathbf{k}|H_m), \quad \theta \in \Theta_n, \quad 1 \leq m \leq M. \quad (3)$$

where $M = \lceil 2^{nR} \rceil$. This is a noncoherent decoding problem because θ is unknown to the receiver. When decoding rule ψ is used, the probability of error is

$$P_e(\theta, \psi) \triangleq \frac{1}{M} \sum_{m=1}^M \Pr[\psi(\mathbf{Y}, \mathbf{K}) \neq m \mid m \text{ sent}, \theta].$$

4.1. Decision Rules

If θ is known to the receiver, the test that minimizes error probability is the maximum likelihood (ML) decision rule

$$\psi_{\text{ML}}(\mathbf{y}, \mathbf{k}) = \operatorname{argmax}_{1 \leq m \leq M} p_\theta(\mathbf{y}, \mathbf{k}|H_m). \quad (4)$$

Denote by $P_e^*(\theta) = P_e(\theta, \psi_{\text{ML}})$ the corresponding error probability; clearly $P_e(\theta, \psi) \geq P_e^*(\theta)$ for any decision rule ψ .

If the receiver does not know θ , there exists generally no decision rule that achieves $P_e^*(\theta)$, i.e., the receiver has to pay a penalty for not knowing θ .

4.2. Asymptotic Optimality

We focus on universal detection rules, which (when they exist) perform as well as the ML detector on the exponential scale. More precisely, a sequence of detection rules ψ_n is said to be universal if

$$\limsup_{n \rightarrow \infty} \max_{\theta \in \Theta_n} \frac{1}{n} \ln \frac{P_e(\theta, \psi_n)}{P_e^*(\theta)} = 0. \quad (5)$$

It should be noted that the frequently used *generalized likelihood ratio test* (GLRT)

$$\psi_{\text{GLRT}}(\mathbf{y}, \mathbf{k}) = \operatorname{argmax}_{1 \leq m \leq M} \max_{\theta \in \Theta_n} p_\theta(\mathbf{y}, \mathbf{k}|H_m) \quad (6)$$

is generally suboptimal. Another popular approach is based on the embedding of pilots, but such an approach is always suboptimal [6].

Universal detection rules exist for some detection problems.

Roughly speaking, the existence of a universal detector depends on the structure of the problem. It is easy to construct examples where a universal detector does not exist. Referring to our attack model for instance, assume $\mathbf{s} = 0$ and

- $T_\theta \mathbf{z} = \theta \mathbf{z}$ and $\Theta_n = \{-1, 1\}$. Therefore $\mathbf{y} = \pm \mathbf{z}$, depending on the value of θ .
- $W(z|x) = \delta(z - x)$ (noiseless channel)
- two hypotheses: channel input is a given sequence \mathbf{x}^0 under H_0 , and sequence $\mathbf{x}^1 = -\mathbf{x}^0$ under H_1 .

When the detector knows θ , it can perfectly discriminate between H_0 and H_1 , and so $P_e^*(\theta) = 0$. When the detector does not know θ , we have complete ambiguity, and $P_e(\theta, \psi) = \frac{1}{2}$ for all ψ . There can be no universal rule for this problem. This example is pathological in that the bipolar signaling scheme used is the worst possible one against the family $\{T_\theta, \theta \in \Theta_n\}$! For any choice of $\mathbf{x}^1 \neq -\mathbf{x}^0$, perfect discrimination is possible, and therefore a universal rule exists.

4.3. Universal Decoding

The above example exhibits the difficulties faced by a particular signal constellation and an attack that exploits a vulnerability of that constellation. Clearly, transmission schemes that are more ‘‘random-like’’ should be less vulnerable to such attacks. This suggests an approach based on the information-theoretic notion of random-coding universality [6], in which the codewords are drawn independently and uniformly from a set \mathcal{B}_n .

We briefly review Feder and Lapidot’s results in [6], in which no side information is present at the transmitter or decoder (no \mathbf{S} , no \mathbf{K}). Let the input sequence \mathbf{x} to the channel be one of the $M = \lceil 2^{nR} \rceil$ elements of a codebook $\mathcal{C} = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^M\}$. These codewords are generated independently according to the uniform distribution on \mathcal{B}_n . The encoder selects a message m and transmits the corresponding codeword \mathbf{x}^m .

Feder and Lapidot studied the optimal-decoding problem for a general family of channels with memory $\{p_\theta(\mathbf{y}|\mathbf{x}), \theta \in \Theta\}$. Under a condition of *strong separability* on the family of channels, they proved the existence of universal decoders in the random coding sense. The universal decoders of [6] are merged list decoders and unfortunately have prohibitive complexity.

As we shall see, the special structure of the family of channels assumed in Sec. 2 simplifies the analysis, even in the presence of side information.

Before concluding this section, observe that under our invertibility assumption **(A1)** on the mappings T_θ , the ML rule based on \mathbf{Y} coincides with the ML rule based on $\mathbf{Z} = U_\theta \mathbf{Y}$. The error probability of the ML decoder is therefore given by

$$\begin{aligned} P_e^*(\theta) &= \frac{1}{M} \sum_{m=1}^M \Pr[\psi_{\text{ML}}(\mathbf{Y}, \mathbf{K}) \neq m \mid m \text{ sent}, \theta] \\ &= \frac{1}{M} \sum_{m=1}^M \Pr[\psi_{\text{ML}}(\mathbf{Z}, \mathbf{K}) \neq m \mid m \text{ sent}] \end{aligned}$$

which is *independent* of θ .

5. NONBLIND WATERMARKING

For nonblind watermarking, $\mathbf{K} = \mathbf{S}$. Let \mathcal{B}_n be the n -dimensional sphere with radius $\sqrt{nD_1}$, where D_1 represents mean-squared watermark embedding distortion. A collection \mathcal{C} of codewords \mathbf{u}^m , $1 \leq m \leq M$ is drawn by sampling from the uniform distribution on \mathcal{B}_n . That is, we use random spherical codes. The marked sequence is given by

$$\mathbf{x} = \mathbf{s} + \mathbf{u}^m.$$

The ML decoding rule (minimize $\|U_\theta \mathbf{y} - \mathbf{s} - \mathbf{u}\|^2$ over $\mathbf{u} \in \mathcal{C}$) is a maximum-correlation rule:

$$\psi_{\text{ML}}(\mathbf{y}, \mathbf{s}) = \operatorname{arg} \max_{1 \leq m \leq M} [\mathbf{u}^m \cdot (U_\theta \mathbf{y} - \mathbf{s})]. \quad (7)$$

The GLRT also takes the form a maximum-correlation rule:

$$\psi_{\text{GLRT}}(\mathbf{y}, \mathbf{s}) = \operatorname{arg} \max_{1 \leq m \leq M} \max_{\theta \in \Theta_n} [\mathbf{u}^m \cdot (U_\theta \mathbf{y} - \mathbf{s}) - \frac{1}{2} \|U_\theta \mathbf{y} - \mathbf{s}\|^2]. \quad (8)$$

Finally, the generalized *normalized correlation* decoder is defined as

$$\psi_{\text{GNC}}(\mathbf{y}, \mathbf{s}) = \operatorname{arg} \max_{1 \leq m \leq M} \max_{\theta \in \Theta_n} \rho(\mathbf{u}^m, U_\theta \mathbf{y} - \mathbf{s}) \quad (9)$$

where

$$\rho(\mathbf{u}, \mathbf{z}) \triangleq \frac{\mathbf{u} \cdot \mathbf{z}}{\|\mathbf{u}\| \|\mathbf{z}\|} \in [-1, 1] \quad (10)$$

is the normalized correlation between sequences \mathbf{x} and \mathbf{z} .

Theorem 5.1 Assume the sequence of index sets Θ_n satisfies Assumption (A2). Then

$$\frac{P_e(\theta, \psi_{\text{GNC}})}{P_e^*(\theta)} \leq |\Theta_n|, \quad \forall \theta \in \Theta_n, \quad (11)$$

and therefore the GNC decoder (9) is universal.

Remark 1. For the GLRT decoder (8), the regret $P_e(\theta, \psi_{\text{GLRT}}) / P_e^*(\theta)$ increases exponentially with n , except when the family of mappings is norm-preserving. Only for this special case is the GLRT decoder is universal.

6. BLIND WATERMARKING

Blind watermarking may be modeled as an information transmission problem with side information \mathbf{S} at transmitter only. Here $\mathcal{K} = \emptyset$. Lattice Quantization Index Modulation is used, specifically the Erez-Zamir random lattice scheme [9], which is capacity-achieving as lattice dimension tends to infinity. Let \mathcal{B}_n be the Voronoi cell for the coarse lattice Λ , with mean-squared embedding distortion D_1 . A collection \mathcal{C} of codewords $\mathbf{u}^m, 1 \leq m \leq M$, is drawn by sampling from the uniform distribution on \mathcal{B}_n .

The transmitter sends

$$\mathbf{x} = \mathbf{s} + [\mathbf{u}^m - \alpha \mathbf{s} - \mathbf{d}] \bmod \Lambda$$

where $\alpha \in (0, 1]$ is the lattice inflation parameter, and \mathbf{d} is an external dither vector, uniformly distributed over \mathcal{B}_n . In the absence of desynchronization attack, the receiver would compute

$$\tilde{\mathbf{y}} = [\alpha \mathbf{y} + \mathbf{d}] \bmod \Lambda$$

and then perform minimum-distance decoding, seeking $\mathbf{u} \in \mathcal{C}$ that minimizes $\|\mathbf{u} - \tilde{\mathbf{y}} \bmod \Lambda\|$.

In the presence of a desynchronization attack, if the receiver knows θ , it applies the inverse mapping U_θ to the received data and applies minimum-distance decoding, i.e., the receiver seeks $\mathbf{u} \in \mathcal{C}$ that minimizes

$$\varphi_{\text{MD}}(\mathbf{u}, \mathbf{y}) = \|\mathbf{u} - \tilde{\mathbf{z}}_\theta \bmod \Lambda\| \quad (12)$$

where

$$\tilde{\mathbf{z}}_\theta \triangleq [\alpha U_\theta \mathbf{y} + \mathbf{d}] \bmod \Lambda.$$

If the receiver does not know θ , it seeks $\mathbf{u} \in \mathcal{C}$ that minimizes

$$\varphi_{\text{GMD}}(\mathbf{u}, \mathbf{y}) = \min_{\theta \in \Theta_n} \|\mathbf{u} - \tilde{\mathbf{z}}_\theta \bmod \Lambda\|. \quad (13)$$

We refer to (12) and (13) as the Minimum-Distance (MD) and the Generalized Minimum-Distance (GMD) decoders, respectively.

Theorem 6.1 Under Assumption (A2), for any $\theta \in \Theta_n$, we have

$$\frac{P_e(\theta, \psi_{\text{GMD}})}{P_e^*(\theta)} \leq |\Theta_n|, \quad (14)$$

i.e., the GMD decoder is universal.

Remark. A similar subexponential bound applies under Assumption (A2').

7. DISCUSSION

From a computational viewpoint, the implementation of the estimator/decoder may be problematic when the parameter set Θ_n is large. There are several practical ideas that could be used to avoid the cost of a full search, and according to the theory presented here, the resulting decoders can be expected to perform very well if an efficient search strategy is found. Also note that the very concept of a full search has been questioned in the watermarking literature, under the belief that a full search would necessarily produce too many false positives. Assumption (A2') suggests some guidelines about discretization of the parameter space Θ_n , in case a full search is computationally feasible.

The work presented in this paper admits several extensions [7]. In one of them, the signals are defined over finite alphabets. In another one, the channel $W(z|x)$ is unknown; and finally, the channel W may have arbitrary memory, subject to almost-sure distortion constraints, as in [11, 12].

8. REFERENCES

- [1] M. Kutter, "Watermarking Resisting to Translation, Rotation and Scaling," *Proc. SPIE*, Boston, Vol. 3528, pp. 423–431, 1998.
- [2] J. J. K. O'Ruanaidh and T. Pun, "Rotation, Scale and Translation Invariant Spread Spectrum Digital Image Watermarking," *Signal Processing*, Vol. 66, No. 3, pp. 303–317, 1998.
- [3] S. Pereira and T. Pun, "Robust Template Matching for Affine Resistant Image Watermarks," *IEEE Trans. on Image Processing*, Vol. 9, No. 6, pp. 1123–1129, June 2000.
- [4] C.-Y. Lin, M. Wu, J. A. Bloom, I. J. Cox, M. L. Miller and Y. M. Lui, "Rotation, Scale, and Translation Resilient Watermarking for Images," *IEEE Trans. on Image Proc.*, Vol. 10, No. 5, pp. 767–782, May 2001.
- [5] M. Costa, "Writing on Dirty Paper," *IEEE Trans. Information Theory*, Vol. 29, No. 3, pp. 439–441, May 1983.
- [6] M. Feder and A. Lapidoth, "Universal Decoding for Channels with Memory," *IEEE Trans. Information Theory*, Vol. 44, No. 5, pp. 1726–1745, Sep. 1998.
- [7] P. Moulin, "Universal Decoding of Watermarks Under Geometric Attacks," *preprint*, May 2006.
- [8] M. Kesal, M. K. Mihçak, R. Kötter and P. Moulin, "Iteratively Decodable Codes for Watermarking Applications," *Proc. 2nd Symposium on Turbo Codes and Related Topics*, Brest, France, Sep. 2000.
- [9] U. Erez and R. Zamir, "Achieving $\frac{1}{2} \log(1 + \text{SNR})$ on the AWGN Channel with Lattice Encoding and Decoding," *IEEE Trans. on Information Theory*, Vol. 50, pp. 2293–2314, Oct. 2004.
- [10] P. Moulin and J. A. O'Sullivan, "Information-Theoretic Analysis of Information Hiding," *IEEE Trans. on Information Theory*, Vol. 49, No. 3, pp. 563–593, 2003.
- [11] A. Somekh-Baruch and N. Merhav, "On the Error Exponent and Capacity Games of Private Watermarking Systems," *IEEE Trans. on IT*, Vol. 49, No.3, pp. 537–562, March 2003.
- [12] P. Moulin and Y. Wang, "On Achievable Error Exponents for Watermarking," *Proc. SPIE Conf.*, San Jose, CA, Jan. 2005.