# On the Capacity of Stegosystems[*]

Pedro Comesaña
Signal Theory and Communications Department
University of Vigo
Vigo, Spain
pcomesan@gts.tsc.uvigo.es

Fernando Pérez-González
Signal Theory and Communications Department
University of Vigo
Vigo, Spain
fperez@gts.tsc.uvigo.es

## ABSTRACT

Among the different applications where data hiding techniques can be used, one that has received huge attention in the last years is steganography. In that scenario, not just the embedded message is hidden, but the communication process itself is tried to be concealed. In spite of the numerous works in this field, the capacity of a perfect stegosystem (meaning a system where it is impossible to know if a given signal is watermarked or not) is still an open question in the data hiding community. In this paper we deal with the capacity of a discrete perfect stegosystem using some optimization procedures, and also present a lower-bound to the capacity of a perfect Gaussian stegosystem; interestingly this bound approaches the capacity of an AWGN channel (without host signal or the perfect steganography constraint) for small (compared with the power of the host) values of the embedding power. Furthermore, we apply the methodology used in this Gaussian scheme to a lattice-based embedding structure, introducing some promising results.

## Categories and Subject Descriptors

K.6.5 [**Management of Computing and Information Systems**]: Security and Protection; H.1.1 [**Models and Principles**]: Systems and Information Theory

## General Terms

Design, Theory

## Keywords

Steganography, Capacity, Costa's Construction, Lattice

## 1. INTRODUCTION

Among the different applications of data hiding techniques which have been studied in the literature, few of them have been paid more attention than steganography. Following Cachin's definition [1] *"steganography is the art and science of communicating in such a way that the presence of a message cannot be detected"*, establishing clearly the difference between the other data hiding applications and steganography. In this application, the target is not only to hide a given message in a host signal, but to conceal the communication process itself. In this way, an unauthorized observer, usually denoted *warden* due to the *prisoners' problem* [13] should not be able to determine if a given signal is watermarked or not.

This condition is usually translated using the Kullback-Leibler distance (KLD) [7], an information theoretic measure which quantifies the similarity between the distributions of two random variables. In fact, Cachin defined that a stegosystem is *perfectly secure* if the KLD between the original host signal and the watermarked one is 0, defending that in that case the warden would not be able to distinguish between both signals [1]. Similarly, Cachin also defined an $\epsilon$-secure stegosystem as that where the former KLD is smaller or equal that $\epsilon$, weakening the initial constraint. At this point we would like to make some remarks on Cachin's nomenclature and properties of the Kullback-Leibler distance:

- Although Cachin denoted those stegosystems where the Kullback-Leibler distance between the original host signal and the watermarked one is 0 as *perfectly secure* (or $\epsilon$-secure when that definition is relaxed), we will refer to that case as *perfectly steganographic stegosystem*, or, for the sake of simplicity, *perfect stegosystem.* With that change of nomenclature we want to emphasize the difference between the steganographic constraint established by Cachin, and the security definitions currently used in the data hiding community [2, 4], related to the impossibility for an attacker of getting knowledge about the secret key. In that sense, Cachin's constraint (which hereafter we will denote as *steganographic constraint*) is more related to the *detectability* of the watermark [11, 12].
- The KLD is not really a distance, as it is not symmetric and does not satisfy the triangle inequality [7].

[*] This work was partially funded by *Xunta de Galicia* under projects PGIDT04 TIC322013PR, PGIDT04 PXIC32202PM, and Competitive research units program Ref. 150/2006; MEC project DIPSTICK, reference TEC2004-02551/TCM and European Comission through the IST Programme under Contract IST-2002-507932 ECRYPT. ECRYPT disclaimer: 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.

*MM&Sec'07,* September 20–21, 2007, Dallas, Texas, USA.

- For discrete random variables, the KLD is always nonnegative and is zero if and only is the two compared distributions coincide.

- For continuous random variables, the KLD is larger or equal to 0, with equality if and only if the two probability density functions (pdf) are equal almost everywhere. In fact, in this paper we will translate the steganographic constraint to the continuous case imposing that both pdfs coincide everywhere.

Based on the former definition by Cachin of a perfectly steganographic system, some works exist in the literature dealing with the search of the capacity of such a system. For example, in [12] Moulin and Wang study the case where the host signal follows a Bernoulli(1/2), both for passive and active wardens (i.e., where the attacker just observes and where he also introduces some attacking noise, respectively). The same authors deal in [14] with the capacity of a perfect stegosystem based on additive spread-spectrum watermarking techniques, and study the KLD obtained when Quantization Index Modulation (QIM) [3] methods are used. Furthermore, the authors present the *Stochastic QIM*, where the original signal is just modified when it lies in the decoding region of a message different from that one wants to embed. Stochastic QIM asymptotically verifies the perfect steganography constraint when the embedding distortion is much smaller than the power of the original host signal, but its capacity is below the capacity achievable by QIM.

Our target in this paper is to quantify how much is lost in terms of achievable rate when the steganographic constraint is verified. To the best our knowledge, this is still an open question; although there are some works in the literature dealing with the loss in the achievable rate in a perfectly steganographic system, as [14], the trade-off between undectability (perfect steganography) and achievable rate has not been measured yet for finite DWRs.

The remainder of the paper is organized as follows: in Sect. 2 we introduce the used notation and framework, which are employed in Sect. 3 and Sect. 4 for studying the capacity of both discrete and continuous stegosystems. The results in the later section are generalized to the case of $\epsilon$-steganographic systems in Sect. 5, whereas Sect. 6 studies a steganographic scheme where lattice-based quantization is used for embedding. Finally, some conclusions are introduced in Sect. 7.

## 2. NOTATION AND FRAMEWORK

In this section we present the notation that will be used throughout the paper. We will denote scalar random variables with capital letters (e.g., $X$) and their outcomes with lowercase letters (e.g. $x$). The same notation criterion applies to random vectors and their outcomes, denoted in this case by bold letters (e.g. $\mathbf{X}$, $\mathbf{x}$). The $i$th component of a vector $\mathbf{X}$ is denoted as $X_i$. In the data hiding problem the embedder wants to transmit a message $m$, which belongs to a $M$-ary alphabet $\mathcal{M} = \{1, \cdots, M\}$, to the decoder sending a signal $\mathbf{y}$ in place of the original host vector $\mathbf{x}$, both of them of length $L$. We will model these signals as realizations of random vectors $\mathbf{Y}$, and $\mathbf{X}$, respectively. The power of the original host signal will be denoted by $\sigma_X^2 \triangleq \frac{1}{L}\sum_{i=1}^{L}\sigma_{X_i}^2$, where $\sigma_{X_i}^2 \triangleq \mathrm{Var}\{X_i\}$, whereas the power of the watermark is given by $D_e \triangleq \frac{1}{L}\sum_{i=1}^{L}\mathrm{E}\{(Y_i - X_i)^2\}$.

Both watermarked and non-watermarked signals are observed by the attacker (warden), who estimates if a given signal was watermarked or not, considering for that purpose the statistical characterization of the source generating the original (non-watermarked) signals. In the case of active attackers a noise vector $\mathbf{n}$ (which can be seen as realization of random vector $\mathbf{N}$, with $\sigma_N^2 \triangleq \frac{1}{L}\sum_{i=1}^{L}\mathrm{E}\{N_i^2\}$) is added to $\mathbf{y}$; therefore, the decoder receives the signal $\mathbf{z} = \mathbf{y} + \mathbf{n}$. Finally, the decoder estimates the embedded symbol with a suitable decoding function. The corresponding block-diagram is plotted in Fig. 1.

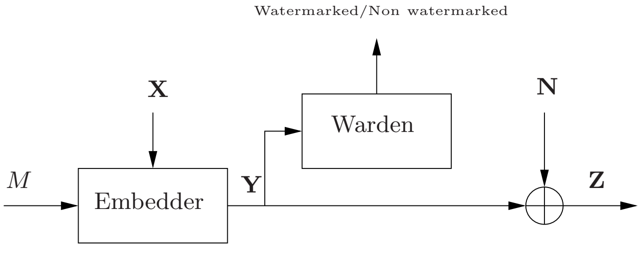

**Figure 1: Block-diagram of a stegosystem.**

For the discrete case the Kullback-Leibler Distance (KLD) is defined as

$$D(p_{\mathbf{X}}||p_{\mathbf{Y}}) = \sum_{\mathbf{x}\in\mathcal{X}} p_{\mathbf{X}}(\mathbf{x})\log\left(\frac{p_{\mathbf{X}}(\mathbf{x})}{p_{\mathbf{Y}}(\mathbf{x})}\right),$$

where $p_{\mathbf{X}}$ and $p_{\mathbf{Y}}$ are the probability mass functions (pmf) of $\mathbf{X}$ and $\mathbf{Y}$ respectively, and $\mathcal{X}$ is the alphabet where both random vectors take values.

On the other hand, for continuous random variables, one has that

$$D(f_{\mathbf{X}}||f_{\mathbf{Y}}) = \int f_{\mathbf{X}}(\mathbf{x})\log\left(\frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{Y}}(\mathbf{x})}\right)d\mathbf{x},$$

where $f_{\mathbf{X}}$ and $f_{\mathbf{Y}}$ denote the pdf of $\mathbf{X}$ and $\mathbf{Y}$, respectively.

In order to compare the power of the host signal and the watermark, we will use the Document to Watermark Ratio (DWR), defined as $\mathrm{DWR} = \frac{\sigma_X^2}{D_e}$; similarly, the Watermark to Noise Ratio (WNR) is defined as $\mathrm{WNR} = \frac{D_e}{\sigma_N^2}$.

## 3. COMPUTATION OF THE CAPACITY OF A DISCRETE STEGOSYSTEM, BASED ON THE JOINT PMF OF $X$, $Y$ AND $U$

As a first step, we consider that the warden is passive, i.e., he is limited to guessing if the observed signal is watermarked or not, but he is not allowed to modify it.

As it is known from Gel'fand and Pinsker's paper [10], the capacity of a system with side information at the embedder is given by

$$C = \max_{p_{U,Y|X}} I(U;Y) - I(U;X),$$

where $U$ is an auxiliar random variable. Taking into account that

$$\begin{aligned} I(U;Y) - I(U;X) &= H(U) - H(U|Y) - H(U) + H(U|X) \\ &= H(U|X) - H(U|Y), \end{aligned}$$

the problem of computing the capacity of a perfect steganographic system can be seen to be equivalent to maximizing the last expression over $p_{U,Y|X}$, constrained to

$$\sum_{u,x} p_{U,Y|X}(u,y|x)p_X(x) = p_Y(y) = p_X(y),$$

due to the steganography constraint, and

$$\sum_{u,y,x} p_{U,Y|X}(u,y|x)p_X(x)(x-y)^2 \leq D_e,$$

the embedding distortion condition.

### 3.1 Upperbounding the capacity with the entropy of the host signal

The introduced framework implies that the capacity of a stegosystem is upperbounded by the entropy of the host signal, because recalling the result by Gel'fand and Pinsker, one can show that

$$\begin{aligned} I(Y;U) - I(X;U) &= H(Y) - H(Y|U) - I(X;U) \\ &\leq H(X), \end{aligned} \qquad (1)$$

where we have used the facts that both $H(Y|U)$ and $I(X;U)$ are non-negative (since we are working with discrete random variables), and $H(Y) = H(X)$, due to the perfect steganographic constraint.

The equality in (1) will be achieved if and only if $H(Y|U) = 0$ and $I(X;U) = 0$; the first condition implies that $Y$ is a function of $U$, while the second one implies that $X$ and $U$ are independent. Taking these circumstances into account, we will try to get some knowledge about how the embedding method should work, or equivalently the structure of $p_{U,Y|X}$, in order to achieve a capacity equal to $H(X)$. In order to do so, we will first show that under the conditions described above, $X$ and $Y$ must be independent:

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(X) - I(X;U|Y) - H(X|U,Y) \\ &= H(X) - H(X) = 0, \end{aligned}$$

where we have used the fact that $I(X;Y|U) = H(Y|U) - H(Y|X,U) = 0 = H(X|U) - H(X|U,Y)$, so $H(X|U,Y) = H(X|U) = H(X)$, and where the last equality is due to the independence of $X$ and $U$. So, once we have shown that $X$ and $Y$ are independent, we can write $p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y) = p_X(x) \cdot p_X(y)$, as the perfect steganography condition establishes that $p_Y = p_X$. Therefore, systems approaching a $H(X)$ capacity must verify

$$\sum_{U} p_{U,X,Y}(u,x,y) = p_X(x) \cdot p_X(y),$$

yielding the following embedding distortion

$$\begin{aligned} D_e^* &= \sum_{x\in\mathcal{X},y\in\mathcal{X}} (x-y)^2 p_X(x)p_X(y) = \sum_{x\in\mathcal{X}} x^2 p_X(x) \\ &+ \sum_{y\in\mathcal{X}} y^2 p_X(y) - 2\left(\sum_{x\in\mathcal{X}} x p_X(x)\right) \cdot \left(\sum_{y\in\mathcal{X}} y p_X(y)\right) \\ &= 2\mathrm{E}\{X^2\} - 2\left(\mathrm{E}\{X\}\right)^2 \\ &= 2\mathrm{Var}\{X\}, \end{aligned}$$

which means that the maximum value of the capacity for any perfect discrete steganographic system, i.e. $H(X)$, is achieved for a DWR of $-3$ dB. Obviously this is not a practical scenario, due to the large embedding distortion that would be introduced, but it is useful for assessing the very limits of a perfectly steganographic system and knowing its asymptotic behavior.

### 3.2 Optimizations results

In order to study the performance of the perfectly steganographic system in more realistic situations, i.e. with larger values of DWR, one must appeal to numerical optimization tools. In this way, in Fig. 2 we have plotted the capacity of a perfectly steganographic system, as well as the capacity of the system with side information at the embedder and taking values only on the alphabet of the host signal, but without the steganographic constraint. For illustration purposes, here we have fixed the pmf of the host signal to a 4 symbols alphabet, following the distribution $p_X(-3) = p_X(3) = 0.05$ and $p_X(-1) = p_X(1) = 0.45$, yielding a host entropy of 1.47 bits. Notice that for medium to high DWRs the differences between the two capacities are negligible, meaning that the steganographic constraint has nearly no impact on the achievable rate. On the other hand, when the DWR is reduced, the capacity of the steganographic system is upperbounded by the maximum capacity given above, i.e. the entropy of the host signal, which in this case takes a value of 1.47 bits, and is achievable, as it was previously said, for a DWR of $-3$ dB. Therefore, one could think of upperbounding the capacity of the stegosystem by the minimum of both the capacity of the side-informed system without the steganographic constraint and the entropy of the host, obtaining a quite good approximation to the real capacity. Another choice for upperbounding the capacity of the stegosystem when the codewords are regularly distributed, as it is in the current case, is the capacity of a lattice-based communication system, meaning a system where the designer has to assign a probability to each centroid of a given lattice (in the current case $2\mathbb{Z}$), trying to maximize the entropy of the resulting signal for a given variance. The advantage of this last alternative is that it is independent of the pmf of the host signal, as long as the the alphabet of the host signal is a subset of the considered lattice.

We can also see that a by-product of the former reasoning is that the cardinality of the alphabet of $U$, $\mathcal{U}$, must verify $|\mathcal{U}| \geq |\mathcal{X}|$ in order to achieve a value of the capacity equal to the entropy of the host signal. In fact, that value of the capacity can be achieved when $|\mathcal{U}| = |\mathcal{X}|$, as we could implement a scheme where the probability of the auxiliar variables $U$ followed the same distribution as $X$, being $X$ and $U$ independent; if we take $Y = U$, it is obvious that the perfect steganographic condition is achieved, and simultaneously the capacity has its maximum value (i.e., $H(X)$), as $Y$ is a function of $U$, and $X$ and $U$ are independent. This strategy can be interpreted as a kind of dither modulation (DM) [3] scheme where each message is related to a one-centroided quantizer.

## 4. COMPUTATION OF A LOWER BOUND ON THE CAPACITY OF A GAUSSIAN PERFECT STEGOSYSTEM

In this section, we introduce some theoretical results on the capacity of a perfect stegosystem based on Costa's construction [6] (therefore, assuming i.i.d. Gaussian host, wa-

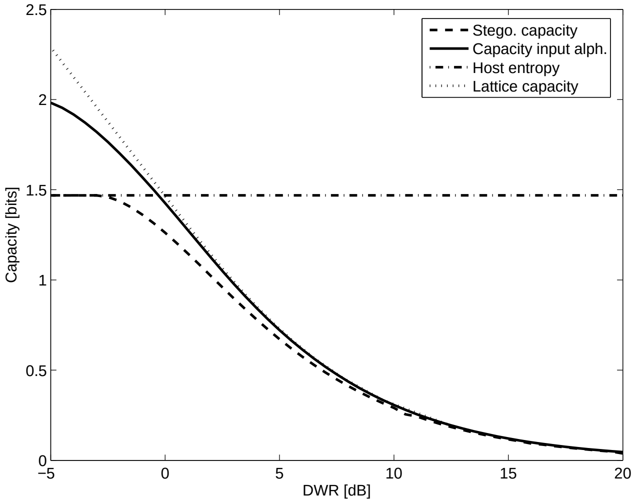

**Figure 2: Comparison of the capacity of the perfectly steganographic system with the capacity of the side-informed scheme using the host alphabet, but without the steganographic constraint. The host entropy and the capacity of the lattice-based communication system are also shown.**

termark and channel noise). This result has to be considered as a lower bound to the capacity of a general steganographic system, as we are fixing its structure without showing it is optimal; nevertheless, the obtained results are asymptotically optimal (in the sense of approaching Costa's result) as the DWR goes to infinity, showing the interest of the obtained bound.

The proposed scheme has the following structure:

$$\begin{aligned} \mathbf{Y} &= \beta(\mathbf{X}+\mathbf{W}), & (2)\\ \mathbf{U} &= \mathbf{W}+\alpha\mathbf{X}, & (3)\\ \mathbf{Z} &= \mathbf{Y}+\mathbf{N}, & (4) \end{aligned}$$

where $\mathbf{W}$ is independent of $\mathbf{X}$, so in order to achieve perfect steganography, $\beta$ must verify $\beta=\sqrt{\frac{\sigma_X^2}{\sigma_X^2+\sigma_W^2}}$. On the other hand, the embedding distortion $D_e$ can be seen to be

$$D_e=\mathrm{E}\{||\mathbf{X}-\mathbf{Y}||^2\}=(1-\beta)^2\sigma_X^2+\beta^2\sigma_W^2,$$

yielding

$$\sigma_W^2=\frac{\sigma_X^2 D_e(4\sigma_X^2-D_e)}{(D_e-2\sigma_X^2)^2}, \qquad (5)$$

which imposes a condition on the range of DWRs where this method could be implemented, as $\sigma_W^2\geq 0$, so $D_e\leq 4\sigma_X^2$, i.e. the DWR has to be greater or equal than $-6$ dB; this condition is exactly the same as that implicitly obtained in [14] for the Spread-Spectrum based scheme. Be aware that this condition will be fulfilled in most practical scenarios.

Considering Gel'fand and Pinsker's result [10], the capacity of this system is given by

$$\begin{aligned} \max I(\mathbf{U};\mathbf{Z})-I(\mathbf{U};\mathbf{X}) &= \max h(\mathbf{Z})-h(\mathbf{Z}|\mathbf{U})\\ &- \; h(\mathbf{X})+h(\mathbf{X}|\mathbf{U}), \qquad (6) \end{aligned}$$

where

$$\begin{aligned} h(\mathbf{Z}) &= \frac{1}{2}\log[2\pi e(\sigma_X^2+\sigma_N^2)],\\ h(\mathbf{X}) &= \frac{1}{2}\log[2\pi e\sigma_X^2]. \end{aligned}$$

On the other hand, for the computation of $h(\mathbf{Z}|\mathbf{U})$, we can write $\mathbf{Z}=c\mathbf{U}+\mathbf{V}_1$, where $\mathbf{V}_1$ is a Gaussian random variable independent of (orthogonal to) $\mathbf{U}$. Considering this decomposition, we can write

$$\mathbf{Z}=c\mathbf{U}+\mathbf{V}_1=c\mathbf{W}+c\alpha\mathbf{X}+\mathbf{V}_1=\beta(\mathbf{X}+\mathbf{W})+\mathbf{N},$$

so

$$\mathbf{V}_1=\mathbf{X}(\beta-c\alpha)+\mathbf{W}(\beta-c)+\mathbf{N},$$

yielding

$$\mathrm{Var}\{\mathbf{V}_1\}=\sigma_X^2(\beta-c\alpha)^2+\sigma_W^2(\beta-c)^2+\sigma_N^2.$$

Furthermore, due to the independence of $\mathbf{U}$ and $\mathbf{V}_1$ we have

$$\begin{aligned} \mathrm{Var}\{\mathbf{Z}\} &= \mathrm{Var}\{c\mathbf{U}\}+\mathrm{Var}\{\mathbf{V}_1\}\\ &= c^2\alpha^2\sigma_X^2+c^2\sigma_W^2+\sigma_X^2(\beta-c\alpha)^2\\ &+ \; \sigma_W^2(\beta-c)^2+\sigma_N^2\\ &= \sigma_X^2+\sigma_N^2, \end{aligned}$$

where the last equality follows from (4) and the perfect steganographic condition. Hence, $c$ must take the value

$$c=\frac{\frac{(D_e-2\sigma_X^2)}{\sigma_X^2}[4\alpha\sigma_X^4+4(1-\alpha)D_e\sigma_X^2-(1-\alpha)D_e^2]}{8\alpha^2\sigma_X^4+8(1-\alpha^2)D_e\sigma_X^2-2(1-\alpha^2)D_e^2},$$

and the variance of $V_1$ is given by

$$\begin{aligned} &\mathrm{Var}\{\mathbf{V}_1\}=\\ &\frac{-(1-\alpha)^2D_e^4-4(1-\alpha)D_e^2\sigma_X^2(-2(1-\alpha)D_e+\sigma_N^2(1+\alpha))}{4\sigma_X^2(4\alpha^2\sigma_X^4+4(1-\alpha^2)D_e\sigma_X^2-(1-\alpha^2)D_e^2}\\ &+\frac{4(1-\alpha)D_e\sigma_X^4(5(-1+\alpha)D_e^2+4(1+\alpha)\sigma_N^2)}{4\sigma_X^2(4\alpha^2\sigma_X^4+4(1-\alpha^2)D_e\sigma_X^2-(1-\alpha^2)D_e^2}\\ &+\frac{16\sigma_X^6((1-\alpha)^2D_e+\alpha^2\sigma_N^2)}{4\sigma_X^2(4\alpha^2\sigma_X^4+4(1-\alpha^2)D_e\sigma_X^2-(1-\alpha^2)D_e^2}. \end{aligned}$$

Similarly, we can write $\mathbf{X}=d\mathbf{U}+\mathbf{V}_2$, with $\mathbf{V}_2$ a Gaussian random variable independent of $\mathbf{U}$:

$$\mathbf{X} \;=\; d\mathbf{U}+\mathbf{V}_2=d\mathbf{W}+d\alpha\mathbf{X}+\mathbf{V}_2,$$

so

$$\mathbf{V}_2=(1-d\alpha)\mathbf{X}-d\mathbf{W},$$

and

$$\mathrm{Var}\{\mathbf{V}_2\}=(1-d\alpha)^2\sigma_X^2+d^2\sigma_W^2.$$

From the independence of $\mathbf{V}_2$ and $\mathbf{U}$, we can write

$$\begin{aligned} \mathrm{Var}\{\mathbf{X}\} &= \mathrm{Var}\{d\mathbf{U}\}+\mathrm{Var}\{\mathbf{V}_2\}\\ &= d^2\sigma_W^2+d^2\alpha^2\sigma_X^2+(1-d\alpha)^2\sigma_X^2+d^2\sigma_W^2\\ &= \sigma_X^2, \end{aligned}$$

implying that

$$d=\frac{\alpha}{\alpha^2+\frac{D_e(4\sigma_X^2-D_e)}{(2\sigma_X^2-D_e)^2}},$$

in such a way that

$$\mathrm{Var}\{\mathbf{V}_2\} = \frac{D_e \sigma_X^2 (4\sigma_X^2 - D_e)}{4\alpha^2\sigma_X^4 + 4(1-\alpha^2)D_e\sigma_X^2 - (1-\alpha^2)D_e^2}.$$

Therefore, (6) can be rewritten as

$$\max I(\mathbf{U};\mathbf{Z}) - I(\mathbf{U};\mathbf{X}) = \max \frac{1}{2}\log\left[\frac{\mathrm{Var}\{\mathbf{Z}\}\mathrm{Var}\{\mathbf{V}_2\}}{\mathrm{Var}\{\mathbf{X}\}\mathrm{Var}\{\mathbf{V}_1\}}\right], \quad (7)$$

which is a function of $\sigma_X^2$, $D_e$, $\sigma_N^2$ and $\alpha$, so the last one is the only parameter we can vary if we want to maximize (7). The value of $\alpha$ that maximizes (7) is denoted by $\alpha^*$ and given by

$$\alpha^* = \frac{D_e(4\sigma_X^2 - D_e)}{4\sigma_X^2(D_e + \sigma_N^2) - D_e^2}, \quad (8)$$

which, when substituted into the rightmost term of (7), yields the following achievable rate, that can be seen as a lower bound for the capacity of a Gaussian stegosystem

$$\frac{1}{2}\log\left(1 + \frac{D_e(4\sigma_X^2 - D_e)}{4\sigma_X^2\sigma_N^2}\right). \quad (9)$$

It is straightforward to see that the steganographic constraint implies a loss in the performance of the presented scheme compared with Costa's original construction [6]; we will quantify such gap with the increase in the WNR needed to achieve a rate equal to Costa's capacity (which is well known to be the capacity of the AWGN channel), obtaining that

$$\begin{aligned}\text{Gap [dB]} &= -10\cdot\log_{10}\left(1 - \frac{D_e}{4\sigma_X^2}\right)\\ &= -10\cdot\log_{10}\left(1 - \frac{1}{4\mathrm{DWR}}\right). \quad (10)\end{aligned}$$

It is particularly interesting to note that when $\sigma_X^2$ goes to infinity, then (10) goes to 0, and (9) approaches Costa's result, i.e.

$$\lim_{\sigma_X^2\to\infty}\frac{1}{2}\log\left(1 + \frac{D_e(4\sigma_X^2 - D_e)}{4\sigma_X^2\sigma_N^2}\right) = \frac{1}{2}\log\left(1 + \frac{D_e}{\sigma_N^2}\right),$$

showing that the proposed scheme is asymptotically optimal from a capacity point of view, being simultaneously perfectly steganographic. Furthermore, under the same conditions, it is straightforward to see that

$$\lim_{\sigma_X^2\to\infty}\alpha^* = \lim_{\sigma_X^2\to\infty}\frac{D_e(4\sigma_X^2 - D_e)}{4\sigma_X^2(D_e + \sigma_N^2) - D_e^2} = \frac{D_e}{D_e + \sigma_N^2},$$

which is nothing but the optimal value of $\alpha$ derived by Costa.

Finally, for comparing this asymptotic behavior with the actual performance when a finite value of $\sigma_X^2$ is used, we have plotted in Fig. 3 the capacity of the proposed scheme as a function of the WNR for different small values of DWR. This plot shows the closeness of the achievable rate to the unconstrained capacity for values of the DWR as small as 5 dB. The distance between the plots for the different studied DWRs is constant for all the range of WNRs, as it was shown in (10), and its value can be observed in Fig. 4. In that figure one can also see the quick decrease of that gap when the DWR is increased; for example, even for small DWRs in practical applications, as it could be the case of a DWR of 20 dB, the gap is as small as 0.01 dB. Furthermore, Fig. 5 plots the value of $\alpha$ used in the studied scenarios, showing again that the larger DWR, the closer the obtained results will be to Costa's values.

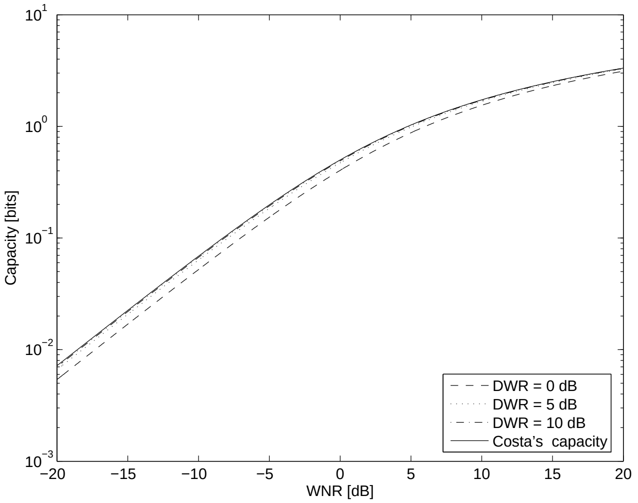

**Figure 3: Comparison of Costa's capacity and the obtained lower bound for different DWRs.**

# 5. $\epsilon$-STEGANOGRAPHIC SYSTEMS

Based on the idea proposed by Cachin in [1] one could weaken the requirement of a steganographic system to be $\epsilon$-steganographic, instead of perfectly steganographic. This would provide the embedder with an additional degree of freedom for increasing the achievable rate, so the expected maximum rate will be larger than in the previous framework. In this section we study this trade-off between the steganographic constraint and the achievable rate, assuming for the sake of simplicity that the watermarked signal is still zero-mean Gaussian, but could have a variance different from that of the original host signal. In this sense, our analysis is giving again a lower bound to the true achievable rate, as for a given value of $\epsilon$, the embedder could also modify the distribution of the watermarked signal, i.e. use a non-Gaussian distribution.

Another question that demands some attention is the order the pdfs appear in that Kullback-Leibler distance; given that the KLD is not symmetric, the obtained result, and therefore the achievable rate for a given $\epsilon$, will be different depending on whether we are computing $D(f_{\mathbf{X}}(\mathbf{x})||f_{\mathbf{Y}}(\mathbf{x}))$ or $D(f_{\mathbf{Y}}(\mathbf{x})||f_{\mathbf{X}}(\mathbf{x}))$. Although the initial proposal of Cachin [1] uses the former, one could also think of using the latter, so we will explore both cases.

## 5.1 Computation of the range of possible variances

Let $f_1(\mathbf{x})$ denote the pdf of a length $L$ Gaussian vector having i.i.d. components with zero mean and variance $\sigma_1^2$. Let $f_2(\mathbf{x})$ be a similar distribution with variance $\sigma_2^2$. In this section we determine the range of values of $\sigma_2^2$ such that the KLD $D(f_1(\mathbf{x})||f_2(\mathbf{x})) \le \epsilon L/2$. This KLD is known to be (e.g., see [14])

$$D(f_1(\mathbf{x})||f_2(\mathbf{x})) = \frac{L}{2}\log\left(\frac{\sigma_2^2}{\sigma_1^2}\right) + \frac{L}{2}\cdot\frac{\sigma_1^2}{\sigma_2^2} - \frac{L}{2}.$$

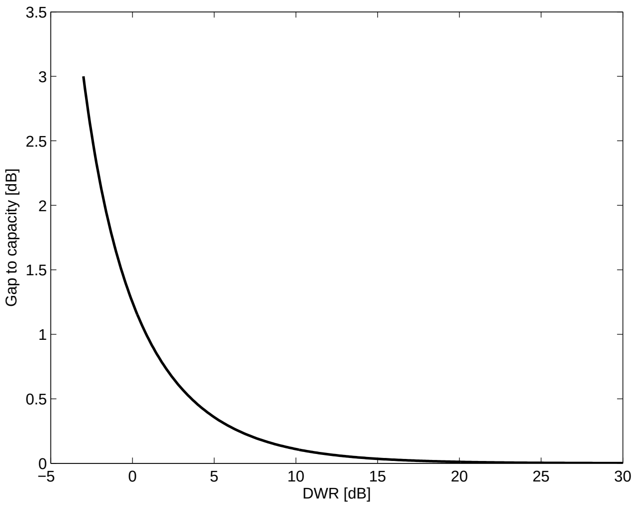

**Figure 4: Increase in the WNR needed to achieve a rate equal to Costa's capacity as a function of the DWR.**

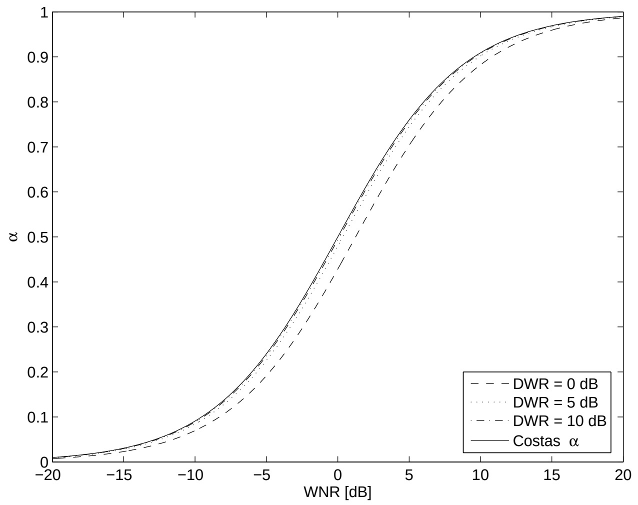

**Figure 5: Comparison of Costa's $\alpha$ and the obtained one for different DWRs.**

In the App. A we show that when $\epsilon > 0$ the function

$$g(x) = -\log(x) + x - 1 - \epsilon, \quad x \geq 0 \tag{11}$$

has exactly two roots, which we denote by $\varphi_1(\epsilon)$ and $\varphi_2(\epsilon)$, with $\varphi_1(\epsilon) < 1 < \varphi_2(\epsilon)$. Furthermore, for any $\epsilon > 0$, $\varphi_1(\epsilon) \cdot \varphi_2(\epsilon) < 1$. Hence, the constraint $D(f_1(\mathbf{x})||f_2(\mathbf{x})) \leq \epsilon L/2$ implies that for a given value of $\sigma_1^2$, the variance $\sigma_2^2$ must verify

$$\sigma_1^2\varphi_1(\epsilon) \leq \sigma_2^2 \leq \sigma_1^2\varphi_2(\epsilon), \tag{12}$$

where $\varphi_1(\epsilon)$ and $\varphi_2(\epsilon)$ are respectively the lower and upper roots of (11). A further characteristic of $\varphi_1(\epsilon)$ and $\varphi_2(\epsilon)$ is that for any $\epsilon > 0$, $\varphi_1(\epsilon)\cdot\varphi_2(\epsilon) < 1$, as it is shown in App. A.

This result just mathematically quantifies how much the variances of two zero-mean Gaussian distributions can differ each other in order to have a KLD smaller or equal than a given value. As it would be intuitively expected, once the variance of one of the Gaussian random vectors is fixed, this automatically defines an interval for the variance of the second random variable. Notice that the variance of the first random variable is always included in such interval, as when the two variances coincide, the KLD is null.

## 5.2 Computation of the maximum achievable rate for a generic power of the watermarked signal

Once we have established the interval of possible variances that will verify the $\epsilon$-steganographic constraint, we have to choose that variance of the watermarked signal which, belonging to this interval, maximizes the achievable rate. In fact, in order to do so, we will use the same strategy followed in Sect. 4, i.e. a scheme based on Costa's construction. Once again, note that the fact that we are dealing with a specific family of distributions, namely Gaussians, and using this particular scheme implies that the obtained achievable rate is just a lower bound to the true capacity of the $\epsilon$-steganographic system.

In the proposed scheme the embedding process is still modeled by equations (2) to (4), so the embedder is free to choose the value of $\beta$ to satisfy the steganographic constraint. If we want the watermarked signal to have a generic power $\sigma_Y^2$, then we have to make $\beta = \sqrt{\frac{\sigma_Y^2}{\sigma_X^2+\sigma_W^2}}$. In that case, the variance of $\mathbf{W}$ is given by

$$\sigma_W^2 = \frac{\sigma_X^2[2D_e(\sigma_X^2+\sigma_Y^2) - D_e^2 - (\sigma_Y^2-\sigma_X^2)^2]}{(\sigma_X^2+\sigma_Y^2-D_e)^2},$$

so, given that $\sigma_W^2 \geq 0$, $D_e$ must verify, $(\sigma_Y - \sigma_X)^2 \leq D_e \leq (\sigma_Y + \sigma_X)^2$.

For computing the maximum achievable rate we will use a procedure similar to that in the previous section, i.e. we will compute the variances of $\mathbf{Z}$, $\mathbf{V}_1$, $\mathbf{X}$ and $\mathbf{V}_2$, and afterwards we will optimize the right hand side of (7).

In order to determine the variance of $\mathbf{V}_1$ we have first to compute the variance of $\mathbf{Z}$, and then compute the parameter $c$ described in the previous section. Following a procedure similar to that in Sect. 4, it is easy to see that

$$\begin{aligned}\mathrm{Var}\{\mathbf{Z}\} &= c^2\alpha^2\sigma_X^2 + c^2\sigma_W^2 + \sigma_X^2(\beta - c\alpha)^2 + \sigma_W^2(\beta-c)^2 \\ &+ \sigma_N^2 = \sigma_Y^2 + \sigma_N^2,\end{aligned}$$

so $c$ will be now given by

$$\begin{aligned}c = {}& \frac{\frac{\sigma_X^2+\sigma_Y^2-D_e}{\sigma_X^2}\left(2\sigma_Y^2[D_e(1-\alpha)+\sigma_X^2(1+\alpha)]\right)}{4[(1-\alpha^2)D_e+\sigma_X^2(1+\alpha^2)] - 2(1-\alpha^2)[(D_e-\sigma_X^2)^2+\sigma_Y^4]} \\ & \frac{\frac{\sigma_X^2+\sigma_Y^2-D_e}{\sigma_X^2}\left(-(1-\alpha)\sigma_Y^4 - (1-\alpha)(D_e-\sigma_X^2)^2\right)}{4[(1-\alpha^2)D_e+\sigma_X^2(1+\alpha^2)] - 2(1-\alpha^2)[(D_e-\sigma_X^2)^2+\sigma_Y^4]},\end{aligned}$$

yielding the following variance of $V_1$

$$\begin{aligned}\mathrm{Var}\{\mathbf{V}_1\} &= [2D_e(\sigma_X^2+\sigma_Y^2) - D_e^2 - (\sigma_Y^2 - \sigma_X^2)^2] \\ &\cdot \ (\sigma_X^2+\sigma_Y^2 - D_e)^2(1-\alpha)^2 \\ &\cdot \ \Bigg(4\sigma_X^2\Bigg(2\sigma_Y^2[D_e(1-\alpha^2)+\sigma_X^2(1+\alpha^2)] \\ &- \ (1-\alpha^2)[(D_e-\sigma_X^2)^2+\sigma_Y^4]\Bigg)\Bigg)^{-1} \\ &+ \ \sigma_N^2.\end{aligned}$$

On the other hand, for the computation of the variance of $\mathbf{X}$ given $\mathbf{U}$, we still have that

$$\mathrm{Var}\{\mathbf{X}\} = d^2\sigma_W^2 + d^2\alpha^2\sigma_X^2 + (1-d\alpha)^2\sigma_X^2 + d^2\sigma_W^2 = \sigma_X^2,$$

so following a reasoning similar to that used in the previous section, $d$ will be now given by

$$d = \frac{\alpha(\sigma_X^2+\sigma_Y^2-D_e)^2}{2[(1-\alpha^2)D_e + \sigma_X^2(1+\alpha^2)] - (1-\alpha^2)[(D_e-\sigma_X^2)^2+\sigma_Y^4]},$$

and the aforementioned variance can be shown to be

$$\mathrm{Var}\{\mathbf{V}_2\} = \frac{\sigma_X^2[2D_e(\sigma_X^2+\sigma_Y^2) - D_e^2 - (\sigma_X^2-\sigma_Y^2)^2]}{2[(1-\alpha^2)D_e + \sigma_X^2(1+\alpha^2)] - (1-\alpha^2)[(D_e-\sigma_X^2)^2+\sigma_Y^4]}.$$

The resulting maximum achievable rate is computed using the rightmost term of (7), and it is maximized when

$$\alpha^* = \frac{D_e^2 + (\sigma_X^2-\sigma_Y^2)^2 - 2D_e(\sigma_X^2+\sigma_Y^2)}{D_e^2 - 4\sigma_N^2\sigma_X^2 + (\sigma_X^2-\sigma_Y^2)^2 - 2D_e(\sigma_X^2+\sigma_Y^2)},$$

taking the following maximum value

$$\frac{1}{2}\log\left(1+\frac{2D_e(\sigma_X^2+\sigma_Y^2) - (\sigma_X^2-\sigma_Y^2)^2 - D_e^2}{4\sigma_N^2\sigma_X^2}\right). \qquad (13)$$

An important observation in (13) is that it is a concave function of $\sigma_Y^2$, achieving its maximum when $\sigma_Y^2 = \sigma_X^2 + D_e$. This implies that when $\sigma_Y^2$ is constrained to take values on a finite interval as that given by (12), the embedder should use $\sigma_Y^2 = \sigma_X^2 + D_e$ when it were feasible, and otherwise the maximum possible value. In the first case $\beta = 1$, so the proposed scheme is reduced to Costa's construction, and the capacity is achieved. On the other hand, when $\epsilon = 0$, i.e. for the perfect stegosystem one has $\sigma_Y^2 = \sigma_X^2$, so the analysis in Sect. 4 can be seen as just a particular case of that made in this section. In the general case, the gap to capacity can be measured as

$$\text{Gap [dB]} = -10\cdot\log_{10}\left(\frac{2(\sigma_X^2+\sigma_Y^2) - \frac{(\sigma_X^2-\sigma_Y^2)^2}{D_e} - D_e}{4\sigma_X^2}\right).$$

## 5.3 Computation of the variance of the watermarked signal

So far we have analyzed what is the relation between the variances of two Gaussian random variables for a given Kullback-Leibler distance, what is the achievable rate for a given $\sigma_Y^2$, and, given a range of possible values of $\sigma_Y^2$, what is value of $\sigma_Y^2$ the embedder will be interested in choosing in order to maximize the achievable rate. Nevertheless, this range of possible values of $\sigma_Y^2$ will depend on the considered KLD, i.e. on whether the constraint is given in terms of $D(f_{\mathbf{X}}(\mathbf{x})||f_{\mathbf{Y}}(\mathbf{x}))$ or $D(f_{\mathbf{Y}}(\mathbf{x})||f_{\mathbf{X}}(\mathbf{x}))$. Next, we discuss the optimal value of $\sigma_Y^2$ for both cases:

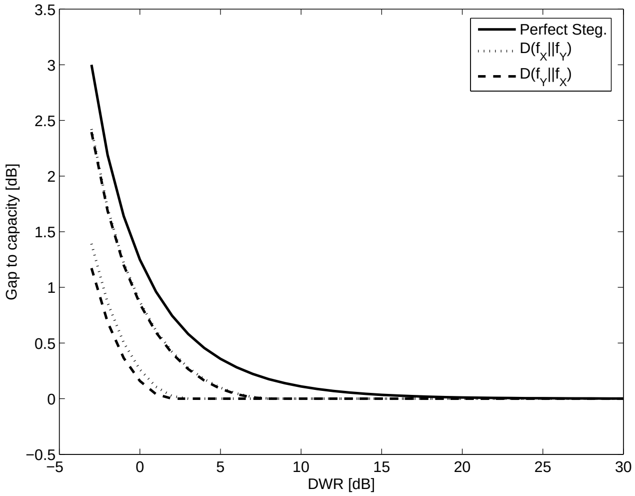

**Figure 6: Increase in the WNR needed to achieve a rate equal to Costa's capacity as a function of the DWR, for the proposed perfect steganographic system, and the $\epsilon$-steganographic ones, for both $D(f_{\mathbf{X}}(\mathbf{x})||f_{\mathbf{Y}}(\mathbf{x}))$ and $D(f_{\mathbf{Y}}(\mathbf{x})||f_{\mathbf{X}}(\mathbf{x}))$. Plots show the results for $\epsilon = 0.1$ and $\epsilon = 0.01$.**

- $D(f_{\mathbf{X}}(\mathbf{x})||f_{\mathbf{Y}}(\mathbf{x})) \leq L\cdot\epsilon/2$: In this case $\sigma_1^2 = \sigma_X^2$ and $\sigma_2^2 = \sigma_Y^2$, so
$$\sigma_Y^2 = \min\{\sigma_X^2\varphi_2(\epsilon), \sigma_X^2 + D_e\}.$$
- $D(f_{\mathbf{Y}}(\mathbf{x})||f_{\mathbf{X}}(\mathbf{x})) \leq L\cdot\epsilon/2$: Now $\sigma_1^2 = \sigma_Y^2$ and $\sigma_2^2 = \sigma_X^2$, yielding that
$$\sigma_Y^2 = \min\left\{\frac{\sigma_X^2}{\varphi_1(\epsilon)}, \sigma_X^2 + D_e\right\}.$$

From the obtained values of $\sigma_Y^2$, and given that $\varphi_1(\epsilon)\cdot\varphi_2(\epsilon) < 1$, a particularly interesting conclusion of this section is that the constraint on $D(f_{\mathbf{Y}}(\mathbf{x})||f_{\mathbf{X}}(\mathbf{x}))$ for a steganographic system will produce more optimistic results, in the sense that the values of $\sigma_Y^2$ will be larger than when the constraint is given in terms of $D(f_{\mathbf{X}}(\mathbf{x})||f_{\mathbf{Y}}(\mathbf{x}))$.

Finally in Fig. 6 one can compare the gap to capacity of the perfect steganographic scheme, and the $\epsilon$-steganographic ones, for both choices of the KLD, i.e. $D(f_{\mathbf{X}}(\mathbf{x})||f_{\mathbf{Y}}(\mathbf{x}))$ and $D(f_{\mathbf{Y}}(\mathbf{x})||f_{\mathbf{X}}(\mathbf{x}))$, for $\epsilon = 0.1$ and $\epsilon = 0.01$. As it was expected, the obtained results for the $\epsilon$-steganographic system with $\epsilon = 0.01$ are closer to those of the perfect steganographic one, than those corresponding to $\epsilon = 0.1$. Another interesting parameter is the point where Costa's capacity can be achieved, i.e., that value of DWR which makes possible to have $\sigma_Y^2 = \sigma_X^2 + D_e$, cancelling the gap, or, in other words, achieving Costa's capacity. When $\epsilon = 0.1$, those values are 2.872 dB and 2.068 dB, while for $\epsilon = 0.01$, 8.293 dB and 8.0730 dB, for $D(f_{\mathbf{X}}(\mathbf{x})||f_{\mathbf{Y}}(\mathbf{x}))$ and $D(f_{\mathbf{Y}}(\mathbf{x})||f_{\mathbf{X}}(\mathbf{x}))$, respectively. These results confirm that the use of $D(f_{\mathbf{Y}}(\mathbf{x})||f_{\mathbf{X}}(\mathbf{x}))$ is somewhat optimistic, compared to $D(f_{\mathbf{X}}(\mathbf{x})||f_{\mathbf{Y}}(\mathbf{x}))$.

# 6. LATTICE-BASED STEGANOGRAPHIC SYSTEMS

Although the results obtained in the two previous sections have a great theoretical interest, as they show how quickly the capacity of a Gaussian steganographic system can approach Costa's capacity when the DWR is increased, their practical application is quite limited, due to the random construction of the codebook and the assumption of a Gaussian-distributed host. In this section we analyze the use of a practical scheme, based on lattice-quantization, for constructing a stego-system with a high achievable rate. This scheme solves the complexity problem raised by the random construction of Costa's scheme. On the other hand, for the sake of tractability, we will maintain our assumption of an i.i.d. Gaussian host.

To the best of authors' knowledge the studied scheme was introduced by the first time by Wang and Moulin in [14], where it is proposed that the output of DC-QIM [3] undergo a postprocessing stage to obtain a watermarked signal with the same power as the original host. However, the results in [14] are only numerical and restricted to the use of uniform scalar quantizers. Both the loss in the achievable rate and the resulting KLD are not analyzed in [14] either.

In this section we propose the use of more sophisticated lattices; specifically, we will use the dirty paper trellis coding scheme introduced by Erez and ten Brink in [8], and applied in [5] to data hiding scenarios. Based on the fact that when the number of dimensions goes to infinity there is a sequence of lattices whose normalized second moment goes to $\frac{1}{2\pi e}$ [9], i.e. whose Voronoi regions tend to hyperspheres, one could think of using those lattices for embedding the information in steganographic schemes. In this way, the distribution of the watermark will resemble the distribution of the Gaussian host. Moreover, recalling the result in [9] that shows that Costa's capacity can be achieved by using lattice-based systems, the embedder will be able to simultaneously increase the achievable rate and reduce the KLD between the original host signal and the watermarked one.

## 6.1 Proposed method

In the *classical* version of DC-QIM, the watermark is constructed as

$$\mathbf{W} = [\mathbf{T} - \alpha\mathbf{X} - \mathbf{D}] \mod \Lambda,$$

where $\Lambda$ is the lattice used for quantizing, $\mathbf{D}$ is a dither vector uniformly distributed over the Voronoi region of $\Lambda$ (usually denoted by $\mathcal{V}(\Lambda)$), $\alpha$ is a scaling parameter, and $\mathbf{T}$ is a vector mapping the message to be embedded. Now, the watermarked signal is given by

$$\mathbf{Y} = \mathbf{X} + \mathbf{W},$$

and the decoder will observe the attacked signal, modeled as

$$\mathbf{Z} = \mathbf{Y} + \mathbf{N},$$

where the noise vector $\mathbf{N}$ is assumed to be i.i.d. Gaussian with variance $\sigma_N^2$.

On the other hand, in the proposal of Wang and Moulin [14] the watermarked signal is given by

$$\mathbf{Y} = \beta(\mathbf{X} + \mathbf{W}),$$

where $\beta = \sqrt{\frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2}}$, and $\sigma_W^2$ is the second moment of the lattice $\Lambda$. Therefore, the embedding power is

$$D_e = (1-\beta)^2\sigma_X^2 + \beta^2\sigma_W^2,$$

so $\sigma_W^2$ can be written as a function of $D_e$ and $\sigma_X^2$,

$$\sigma_W^2 = \frac{\sigma_X^2 D_e(4\sigma_X^2 - D_e)}{(D_e - 2\sigma_X^2)^2}, \tag{14}$$

coinciding with the value obtained in (5) for the perfectly steganographic system based on Costa's construction; as in that case, the proposed analysis only makes sense when DWR $\geq -6$ dB.

One question that was not addressed in [14], but which needs to be solved is how the received signal is processed before performing the decoding, as the scaling by $\beta$ would be modifying the considered codebook. In this paper we will use the most immediate procedure, that amounts to rescaling the received signal. Nevertheless, we would like to remark that this may be a non-optimal strategy if one wants to maximize the achievable rate. Hence, we will analyze the case where the decoding is based on the random variable[1]

$$\mathbf{Z}' = \left[\frac{\alpha\mathbf{Z}}{\beta} + \mathbf{D}\right] \mod \Lambda. \tag{15}$$

## 6.2 Achievable rate analysis

In order to compute the achievable rate, we need to characterize the random variable $\mathbf{Z}'$. From (15) we can see that

$$\begin{aligned}
\mathbf{Z}' &= \left[\alpha\mathbf{X} + \alpha\mathbf{W} + \frac{\alpha\mathbf{N}}{\beta} + \mathbf{D}\right] \mod \Lambda \\
&= \left[\mathbf{T} - (1-\alpha)\mathbf{W} + \frac{\alpha\mathbf{N}}{\beta}\right] \mod \Lambda \\
&= \left[\mathbf{T} + \mathbf{N}'\right] \mod \Lambda,
\end{aligned}$$

where $\mathbf{N}'$ is defined as

$$\mathbf{N}' \triangleq \left[-(1-\alpha)\mathbf{W} + \frac{\alpha\mathbf{N}}{\beta}\right] \mod \Lambda.$$

Given that the achievable rate is given by

$$\frac{1}{L}I(\mathbf{T};\mathbf{Z}') = \frac{1}{L}h(\mathbf{Z}') - \frac{1}{L}h(\mathbf{Z}'|\mathbf{T}), \tag{16}$$

we can follow the reasoning in [9] to establish that the achievable rate is maximized when $T \sim U(\mathcal{V}(\Lambda))$. On the other hand, we can bound

$$\begin{aligned}
\frac{1}{L}h(\mathbf{Z}'|\mathbf{T}) &= \frac{1}{L}h(\mathbf{N}') \leq \frac{1}{L}h\left(-(1-\alpha)\mathbf{W} + \frac{\alpha\mathbf{N}}{\beta}\right) \\
&\leq \frac{1}{2}\log\left(2\pi e\left[(1-\alpha)^2\sigma_W^2 + \frac{\alpha^2\sigma_N^2}{\beta^2}\right]\right).
\end{aligned}$$

With the previous considerations in mind, the achievable rate can be bounded like

$$\begin{aligned}
\frac{1}{L}I(\mathbf{T};\mathbf{Z}') &\geq \frac{1}{2}\log\left(\frac{\sigma_W^2}{G(\Lambda)}\right) \\
&- \frac{1}{2}\log\left(2\pi e\left[(1-\alpha)^2\sigma_W^2 + \frac{\alpha^2\sigma_N^2}{\beta^2}\right]\right). 
\end{aligned} \tag{17}$$

[1] This choice is equivalent to making the decision based on the variable $[\alpha\mathbf{Z} + \beta D] \mod (\beta\Lambda)$.

where we have used the fact that $\mathbf{Z}'$ is uniformly distributed over $\mathcal{V}(\Lambda)$. Therefore, the embedder will be interested in looking for the value of $\alpha$ that maximizes (17), which after replacing $\sigma_W^2$ in (17) by the rightmost term of (14) can be shown to be

$$\alpha^* = \frac{D_e(4\sigma_X^2 - D_e)}{4\sigma_X^2(D_e + \sigma_N^2) - D_e^2}, \tag{18}$$

which is the same value obtained in Sect. 4 when the random codebook based stegosystem was analyzed. Replacing $\alpha^*$ in (17), one obtains

$$\frac{1}{L} I(\mathbf{T};\mathbf{Z}') \geq \frac{1}{2}\log\left(1 + \frac{D_e(4\sigma_X^2 - D_e)}{4\sigma_X^2\sigma_N^2}\right) - \frac{1}{2}\log(2\pi e G(\Lambda)).$$

Finally, as it was previously said, there is a sequence of lattices such that $G(\Lambda) \to \frac{1}{2\pi e}$ when the number of dimensions goes to infinity, so we can lower-bound the achievable rate of the proposed system by

$$\frac{1}{L} I(\mathbf{T};\mathbf{Z}') \geq \frac{1}{2}\log\left(1 + \frac{D_e(4\sigma_X^2 - D_e)}{4\sigma_X^2\sigma_N^2}\right). \tag{19}$$

It is remarkable that the maximum achievable rate in this case coincides with that obtained in Sect. 4, so the conclusions extracted there are still valid in this framework, including our comments regarding the gap to capacity. Therefore, when the DWR goes to infinity, the rightmost term of (19) approaches Costa's capacity, implying that the proposed scheme is asymptotically optimal. The behavior of the achievable rate, the gap to capacity, and the value of $\alpha^*$ can be checked for finite values of DWR in Fig. 3-5, respectively, as their values coincide with the computed ones for Costa's construction based scheme, when the dimensionality of the problem goes to infinity.

This result shows that the equivalence in the performance of random codebooks and lattice-based schemes (with lattices going to hyperspheres) is not only applicable to data hiding capacity computation, but also to the computation of the maximum achievable rate of stegosystems.

### 6.3 Analysis of the steganographic constraint

So far we have shown, from the point of view of the achievable rate, the advantages of using lattices whose Voronoi regions tend to hyperspheres. In this section we show the advantages of those lattices also for verifying the steganographic constraint. The intuitive idea is that a random variable uniformly distributed over a hypersphere, as for example the watermark obtained using the mentioned lattices, tends assymptotically to a Gaussian distribution when the dimensionality goes to infinity. Therefore, it is expected that the more similar the Voronoi region of the lattice is to a hypersphere, the lower the KLD between the host and the watermarked signals will be.

In Fig. 7 we can see $D(f_{\mathbf{X}}(\mathbf{x})||f_{\mathbf{Y}}(\mathbf{x}))$ for three differents schemes: the basic DC-DM based on uniform scalar quantizers without scaling, DC-DM also based on uniform scalar quantizers but scaling by $\beta$, and the method by Erez and ten Brink [8] also with scaling. In this last case one can observe a floor in the KLD; this floor is due to the empirical computation of the KLD for this scheme. Nevertheless, for the range of DWRs where that floor is not present, the KLD obtained for Erez and ten Brink's lattice is much lower (about one order of magnitude) than the one obtained for uniform scalar DC-DM with scaling. Finally, as expected, the results obtained for DC-DM without scaling are worse that those obtained with scaling.

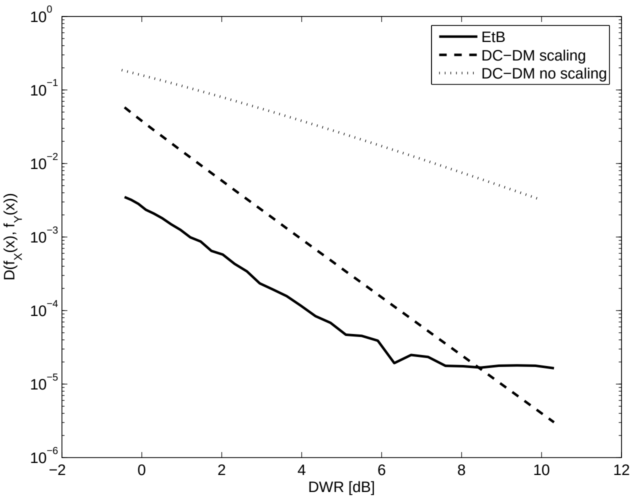

**Figure 7: $D(f_{\mathbf{X}}(\mathbf{x})||f_{\mathbf{Y}}(\mathbf{x}))$ obtained using Erez and ten Brink's scheme scaled by $\beta$, DC-DM using uniform scalar quantizers also scaled by $\beta$, and DC-DM using uniform scalar quantizers without scaling.**

In the final version of the paper, we will introduce theoretical approximations to these plots.

## 7. CONCLUSIONS

In this paper we have tried to shed some light on the open question of stegosystems capacity. Accurate lower bounds have been provided, both for the discrete and for the continuous Gaussian case. Some of the main conclusions we obtained are the following:

- For the discrete case, the capacity of the stegosystem is bounded by the entropy of the host signal. The capacity can take that value for values of DWR smaller or equal than $-3$ dB. Accurate approximations to the real capacity were provided.

- In the Gaussian case we proposed a stegosystem based on Costa's construction, which is shown to provide an achievable rate very close to Costa's capacity, being this gap reduced when the DWR is increased.

- The maximum achievable rate for the $\epsilon$-steganographic version of the previous scheme is also analyzed. As it was expected, the obtained values in that case are closer to Costa's capacity, as the embedder will have an additional degree of freedom.

- Finally, the use of lattice-based data hiding is shown to be a good choice for performing steganography, specially when one works with lattices whose Voronoi regions tend to hyperspheres. In this case, the achievable rate is maximized, coinciding with that obtained by Costa's based construction, and simultaneously the KLD between the original host signal and the watermarked one seems to be minimized.

## 8. REFERENCES

[1] C. Cachin. An information-theoretic model for steganography. In D. Aucsmith, editor, *Information Hiding International Workshop*, volume 1525 of *Lecture Notes in Computer Science*, pages 306–318, Portland, OR, USA, April 1998. Springer.

[2] F. Cayre, C. Fontaine, and T. Furon. Watermarking security: theory and practice. *IEEE Transactions on Signal Processing*, 53(10):3976–3987, October 2005.

[3] B. Chen and 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):1423–1443, May 2001.

[4] P. Comesaña, L. Pérez-Freire, and F. Pérez-González. An information-theoretic framework for assessing security in practical watermarking and data hiding scenarios. In *6th International Workshop on Image Analysis for Multimedia Interactive Services*, Montreux, Switzerland, April 2005.

[5] P. Comesaña, F. Pérez-González, and F. M. J. Willems. Applying Erez and ten Brink's dirty paper codes to data-hiding. In E. J. Delp III and P. W. Wong, editors, *Proceedings of SPIE*, volume 5681 of *Security, Steganography and Watermarking of Multimedia contents VII*, pages 298–307, San Jose, CA, USA, January 2005. SPIE.

[6] M. H. M. Costa. Writing on dirty paper. *IEEE Transactions on Information Theory*, 29(3):439–441, May 1983.

[7] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley series in Telecommunications, 1991.

[8] U. Erez and S. ten Brink. A close-to-capacity dirty paper coding scheme. *IEEE Transactions on Information Theory*, 51(10):3417–3432, October 2005.

[9] U. Erez and R. Zamir. Achieving $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding. *IEEE Transactions on Information Theory*, 50(10):2293–2314, October 2004.

[10] S. I. Gel'fand and M. S. Pinsker. Coding for channel with random parameters. *Problems of Control and Information Theory*, 9(1):19–31, 1980.

[11] P. Moulin and R. Koetter. Data-hiding codes. *Proceedings of the IEEE*, 93(12):2083–2126, December 2005.

[12] P. Moulin and Y. Wang. New results on steganographic capacity. In *Proceeding CISS Conference*, Princeton, NJ, March 2004.

[13] G. J. Simmons. The prisoners' problem and the subliminal channel. In *Advances in Cryptology: CRYPTO83*, pages 51–67, August 1984.

[14] Y. Wang and P. Moulin. Steganalysis of block-structured stegotext. In Edward J. Delp III and Ping W. Wong, editor, *Security, Steganography, and Watermarking Multimedia Contents VI*, volume 5306, pages 477–488, San Jose, CA, 2004. SPIE.

## APPENDIX

## A. PROOF OF THE PROPERTIES OF $\varphi_1(\epsilon)$ AND $\varphi_2(\epsilon)$.

Given the continuous and concave nature of $g(x)$, the fact that it is not bounded, and that it achieves its minimum at $x=1$, where it is equal to $-\epsilon$, $g(x)$ has just a root in $x=1$ if $\epsilon=0$ (implying $\sigma_1^2=\sigma_2^2$, as it was expected), and two roots, one larger and other smaller than 1, when $\epsilon>0$.

On the other hand, when $\epsilon>0$, given that $g(x)$ is strictly increasing for $x\in(1,\infty)$, the inequality $\varphi_1(\epsilon)\cdot\varphi_2(\epsilon)<1$, can be seen to be equivalent to

$$g\left(\frac{1}{\varphi_1(\epsilon)}\right)>0,$$

or, equivalently

$$\varphi_1(\epsilon)+\frac{1}{\varphi_1(\epsilon)}-2(1+\epsilon)>0, \tag{20}$$

where we have used the fact that $\log(\varphi_1(\epsilon))=\varphi_1(\epsilon)-1-\epsilon$. In order to prove (20) we will follow the *Reductio ad absurdum* argument, assuming that $g\left(\frac{1}{\varphi_1(\epsilon)}\right)\leq 0$, and showing then that it is not possible.

If $g\left(\frac{1}{\varphi_1(\epsilon)}\right)\leq 0$, then from (20) it must be verified that $\varphi_1(\epsilon)\in[1+\epsilon-\sqrt{2\epsilon+\epsilon^2},1+\epsilon+\sqrt{2\epsilon+\epsilon^2}]$. Given that by definition $\varphi_1(\epsilon)<1$, then the previous range can be reduced to $[1+\epsilon-\sqrt{2\epsilon+\epsilon^2},1)$. If one computes $g(1+\epsilon-\sqrt{2\epsilon+\epsilon^2})$ is easy to show that it is negative (as it is 0 for $\epsilon=0$, and its derivative with respect to $\epsilon$ is negative). From the fact that $g(x)$ is strictly decreasing for $x\in(0,1)$, it is straightforward to see that $g(x)<0$ for any $x\in[1+\epsilon-\sqrt{2\epsilon+\epsilon^2},1)$, so it is not possible that $\varphi_1(\epsilon)$ belongs to that interval.