Security of Lattice-Based Data Hiding Against the Known Message Attack

Luis Pérez-Freire*, Fernando Pérez-González, Teddy Furon, Pedro Comesaña

Abstract

Security of Quantization Index Modulation (QIM) watermarking methods is usually sought through a pseudorandom dither signal which randomizes the codebook. This dither plays the role of the secret key, i.e. a parameter only shared by the watermarking embedder and decoder, which prevents unauthorized embedding and/or decoding. However, if the same dither signal is reused, the observation of several watermarked signals can provide sufficient information for an attacker to estimate the dither signal. This paper focuses on the cases when the embedded messages are either known or constant. In the first part of this paper, a theoretical security analysis of QIM data hiding measures the information leakage about the secret dither as the mutual information between the dither and the watermarked signals. In the second part, we show how set-membership estimation techniques successfully provide accurate estimates of the dither from observed watermarked signals. The conclusion of this twofold study is that current QIM watermarking schemes have a relative low security level against this scenario because a small number of observed watermarked signals yield a sufficiently accurate estimate of the secret dither. The analysis presented in this paper also serves as the basis for more involved scenarios.

Index Terms

Watermarking security, Quantization Index Modulation, lattice data hiding, mutual information, equivocation, set-membership estimation.

I. Introduction

Recently, the basis of cryptanalysis has been cast to data hiding to establish the concept of watermarking security [1], [2], [3]. It assumes that all details of the watermarking technique are publicly known except the so-called secret key parameter of the embedding and decoding processes, according to Kerckhoff's principle [4]. Hence, security only

Luis Pérez-Freire, Fernando Pérez-González and Pedro Comesaña are with the Signal Theory and Communications Department, ETSI Telecom., University of Vigo, 36310 Vigo, Spain (e-mail: {lpfreire,fperez,pcomesan}@gts.tsc.uvigo.es)

Teddy Furon is with IRISA/TEMICS, Campus Universitaire de Beaulieu, 35042 Rennes Cedex, France (e-mail: teddy.furon@irisa.fr).

This work was partially funded by *Xunta de Galicia* under projects PGIDT04 TIC322013PR and PGIDT04 PXIC32202PM; MEC project DIPSTICK, reference TEC2004-02551/TCM; FIS project IM3, reference G03/185, European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT, and Fundación Caixa Galicia grant for postgraduate studies. 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.

relies on whether (or more realistically, for how long) the secret key will remain secret. This framework for security assessment of watermarking schemes is twofold. We assume that a collection of content has been watermarked with the same secret key, and the attacker has access to these watermarked signals. A first theoretical part measures the amount of information about the secret key which leaks from the watermarked content, using the mutual information and conditional entropy as measures, following the information-theoretic approach for cryptosystems proposed by Shannon in [5]: this approach is based on computing the entropy of the key conditioned on the encrypted messages observed by the attacker; when the conditional entropy is null it means that the attacker has gathered enough observations so as to disclose the secret key. Bear in mind that the original Shannon's work dealt with discrete random variables, whereas our discussion deals with continuous random variables; this is why we need to resort to differential entropies, nevertheless the main concepts remain the same: whereas discrete entropy is related to the number of possible values of a random variable and their probabilities (and it is always a non-negative quantity), differential entropy accounts for the log-volume of the typical set [6, Section 9.2] and as such still provides a useful measure of uncertainty, regardless of whether it takes negative values; in particular, complete disclosure of the secret key will be possible when its conditional entropy becomes $-\infty$. The information-theoretic analysis allows us to establish lower bounds on the variance of the key estimation error as a function of the number of available observations. The second part of the paper is of practical nature and shows workable algorithms which take as input a collection of watermarked signals and output an estimate of the secret key. This confirms that the attack is manageable within a bounded complexity.

This framework (theoretical and practical parts) has already been successfully applied to substitutive [2, Sect. III] and additive spread spectrum watermarking schemes [2, Sect. IV],[3]. Watermarking security under this viewpoint is also briefly addressed in Section 10 of [7]. Other notable works dealing with the security of spread spectrum schemes concentrate on the practical part [8], [9]. As for quantization based data hiding, preliminary studies of the theoretical part have shown the existence of information leakages [10], while on the practical part, we are aware of two works: first, the work by J. Eggers *et al.* [11], although their motivation was not the security analysis but the robustness improvement of the Scalar Costa Scheme (SCS) against a Scaling and Addition of White Gaussian Noise attack (SAWGN), and the work by Bas and Hurri [12] which is more related to our approach, but focuses on the so-called spread transform methods [13] without distortion compensation.

The reader must note that the scope of this article is restricted because we mainly focus on the Known Message Attack (KMA) [2, Sect. II.B]: we assume that the attacker is able to gather a collection of signals $\{\mathbf{y}_i\},\ i = 1, \ldots, N_o$, watermarked with the same key, while knowing for each its hidden message, denoted by m_i. The pairs $\{\mathbf{y}_i, m_i\}$ will be referred to as *observations*. This paper is only a first step to a global security analysis of quantization-based watermarking; in fact, as we will discuss in Section VII, both the developed theory and algorithms for the KMA scenario constitute the core of those corresponding to more complex scenarios. In any case, the considered setup is still very important as shown in the following motivations.

1) The copy protection application faces extreme security threats [14]. The secret key is not only unique but the hidden messages are also known by any user. Consider a *Digital Rights Management* (DRM) system using watermarking. The secret key is embedded in a chipset included in every compliant device. Content makers also

share the secret key to watermark their products. Compliant devices spot these as protected contents whose usage is restricted according to the DRM license. Some DRM systems hide the status (i.e., the usage restriction) such as ‘Copy Never’, ‘Copy Once’, ‘Copy No More’ [14] in the contents. The number of status choices is extremely small compared to the size of the content. This is a typical example of zero-rate watermarking, where the embedding proceeds by blocks (of video or of sound). The *Copy Protection Technical Working Group* has, for instance, required the embedding of eight bits within ten seconds of video [15]. Moreover, any user knows the embedded message as the status of a piece of content is public (for instance, the compliant device may warn the user that the copy of a particular content is forbidden due to its restrictive status). Hence, KMA is a main threat in copy protection applications.

2) Video and audio watermarking in general might be also put at risk by KMA. The reason is that one usually does not watermark a video, but instead watermarks consecutive blocks of video. This division in blocks maintains a low complexity of the embedding and decoding whereas it eases temporal re-synchronization. In the case of zero-rate watermarking (the message space is bounded and small), a common approach is to embed the message repeatedly in consecutive blocks. The division into blocks is usually publicly known (although other strategies are possible), so the attacker is able to gather a number of different blocks hiding the same message. This is not exactly a KMA but a Constant Message Attack (CMA) because the value of the message might not be known. However, we will show that this only brings slight changes in both the theoretical and the practical parts. This matter concerns applications such as copyright enforcement, copy protection, and fingerprinting (traitor tracing). Note that in this last scenario the major source of concern has been collusion attacks (i.e., an arrangement of several traitors), but the CMA could constitute a worse attack for audio or video fingerprinting because there the same message is repeatedly embedded in a block by block basis. The success of the CMA depends on the number of blocks in a movie or song.

3) Another motivation is that most QIM schemes are known to be weak against amplitude scaling attacks. Eggers *et al.* suggest in [11], [16, Sect. VI] to embed a reference message (aka pilot sequence) prior to the message to be embedded. Knowing this pilot sequence, the decoder is able to estimate the amplitude scaling and later to retrieve the hidden message. Once again, this implies that all the watermarked signals contain the same pilot sequence. If the CMA is successful, the attacker may remove the pilot sequence and then apply a slight scaling to the amplitude of the host signal. The decoder will not be able to retrieve the reference message nor the scaling factor. However, bear in mind that in case the attacker does not know the exact location of the pilot signal, the CMA attack can be used only as part of a more global attack.

4) Another important case is the application of QIM schemes to authentication. There exist many different ways of designing a watermarking-based authentication scheme; for instance, Eggers *et al.* proposed a highly original scheme [17] taking benefit that a side-informed embedder gives a watermark signal heavily dependent on the host signal. Thus, it is useless to estimate the watermark signal of a signed content, and then copy and paste it into another content in order to forge a signature. This elegantly gets rid of the copy and paste attack. In their scheme, Eggers *et al.* suggest to apply SCS to image authentication by watermarking blocks of the image with a reference message. The verification process considers the image as authentic when the decoded message matches this reference message. Once again, this implies that all signed images contain the same reference message and the CMA attack is applicable.

We assume the watermark embedder works as follows. A first step extracts some coefficients (DCT, DWT, FFT...) from the original piece of content. These coefficients are ordered in a length-N_v column vector, and the latter is partitioned in l blocks of length n, denoted by $\mathbf{x}_i,\ i = 1, \ldots, l-1$. The embedder hides a message $m_i \in \mathcal{M}$ in each $\mathbf{x}_i$, yielding a watermarked vector $\mathbf{y}_i$. Thus, the data hiding rate is $R = \log_2(|\mathcal{M}|)/n$ bits per coefficient. The specific implementation of QIM considered in this paper is by means of nested lattices [7], which encompasses most of the proposed QIM formulations so far, and it will be referred to as *lattice data hiding*. According to the discussion above, we assume that both the selection of the extracted coefficients and the partitioning in length-n blocks is public; hence, the security of the scheme relies only in the randomization of the lattice via a dithering process, where the dither signal plays the role of secret key.[1] Actually, the secret dither signal, which we denote by $\mathbf{t}$, may be any deterministic function of a certain cryptographic key $\boldsymbol{\theta}$, i.e., $\mathbf{t} = g(\boldsymbol{\theta})$, where $g(\cdot)$ is a pseudo-random generator. Although, under the assumptions of Kerckhoff's principle, such function should be publicly known, disclosure of the secret dither does not necessarily imply disclosure of the cryptographic key $\boldsymbol{\theta}$, since $g(\cdot)$ should have been properly designed so as to be (ideally) non-invertible. The attacker restricted to a signal processing approach, as the one we are presenting here, can, at most, aspire to disclose the sequence of dither samples provided by the available observations. Inference of the secret key based on the estimate of $\mathbf{t}$ belongs to the domain of cryptanalysis, and as such falls out of the scope of the present paper. Nevertheless, the mere disclosure of the plain secret dither in a lattice data hiding scheme allows many harmful attacks, as we shall discuss in Section V-A.

The theoretical security of lattice data hiding schemes is studied in sections II and III. Sections IV and V present practical estimators and experimental results, respectively, obtained in the lattice data hiding scenario. In Section VI, the theoretical security of quantization-based data hiding methods is linked to the corresponding to Costa's set-up [18], and in Section VII the extension of the framework proposed in this paper to more general scenarios is discussed. Finally, in Section VIII the main conclusions are summarized and some remarks are given. Unless otherwise stated, our results will be restricted to a distortion compensation parameter $\alpha \geq 0.5$ which represents the most important case for lattice data hiding for the following reasons:

- In high Watermark to Noise Ratio (WNR) applications,[2] which is the scenario of main interest for lattice data hiding, the optimal value of α is considerably larger than 0.5 (see [16], for instance).
- In low WNR applications, the optimal values of α are smaller than 0.5, leading to decoding errors even in the absence of noise. Indeed, it has been shown that for low WNR's it is better to apply lattice data hiding in conjunction with spread transform [13], whose main benefit is to increase the effective WNR. This in turn leads to an increase of the optimal α, in most practical instances to values ≥ 0.5. A similar conclusion is arrived at when lattice data hiding is combined with channel coding (e.g., repetition coding or Construction A [19]).

The main notational conventions followed throughout the text are the following: random variables and their occurrences are denoted by capital and lowercase letters, respectively; boldface letters denote column vectors, whereas

[1]The extension to more general scenarios using secret coefficient permutations, for instance, will be addressed in Section VII.

[2]WNR $\triangleq \log_{10}(D_w/\sigma_N^2)$, where D_w and σ_N^2 are the the embedding distortion and noise power per dimension, respectively. Throughout this paper, the terms high and low WNR are loosely applied to WNR > 0 dB and WNR ≤ 0 dB, respectively.

scalar variables are represented in non-boldface characters. Calligraphic letters are reserved for sets. All logarithms are to the base e, so all the mutual informations and differential entropies are expressed in natural units.

II. THEORETICAL SECURITY OF LATTICE-BASED DATA HIDING

Before proceeding with the theoretical analysis, we will briefly explain the basics of embedding in lattice data hiding; for more details and other aspects such as decoding, the interested reader is referred to [7] and the references therein. Consider an n-dimensional lattice Λ and the set $\mathcal{M} = \{0, \ldots, L_M - 1\}$ of possible messages. For each message $m \in \mathcal{M}$ let us define the associated coset of Λ as $\mathcal{U}_m \triangleq \Lambda + \mathbf{d}_m$, where $\mathbf{d}_m$ is the minimum-norm *coset representative* corresponding to message m. The codebook $\mathcal{U}$ is defined by the union of all cosets, $\mathcal{U} \triangleq \bigcup_{m=0}^{L_M-1} \mathcal{U}_m$. Given a certain host signal $\mathbf{x}$ and a to-be-transmitted message $m \in \mathcal{M}$, each watermarked signal is generated as

$$\mathbf{y} \;\; = \;\; \mathbf{x} + \alpha \left(Q_{\mathcal{U}_m}(\mathbf{x}) - \mathbf{x} \right), \tag{1}$$

where α is the distortion compensation parameter, and $Q_{\mathcal{U}_m}(\cdot)$ is an Euclidean quantizer whose centroids are defined by the coset $\mathcal{U}_m$:

$$Q_{\mathcal{U}_m}(\mathbf{x}) \triangleq \arg \min_{\mathbf{r} \in \mathcal{U}_m} ||\mathbf{x} - \mathbf{r}||, \tag{2}$$

where $||\cdot||$ denotes Euclidean norm. This data hiding scheme is commonly known as Distortion Compensated - Dither Modulation (DC-DM) [13]. For adding security to the scheme, several authors [13],[16] proposed to introduce an additional term $\mathbf{t}$ named *secret dither vector*, which is known only by embedder and decoder, yielding a randomized embedding function:

$$\mathbf{y} = \mathbf{x} + \alpha(Q_{\mathcal{U}_m,\mathbf{t}}(\mathbf{x}) - \mathbf{x}) = Q_{\mathcal{U}_m,\mathbf{t}}(\mathbf{x}) + (1-\alpha)(\mathbf{x} - Q_{\mathcal{U}_m,\mathbf{t}}(\mathbf{x})), \tag{3}$$

where $\mathcal{U}_{m,\mathbf{t}} \triangleq \Lambda + \mathbf{d}_m + \mathbf{t}$ is the m-th randomized coset, and the second term of (3) is the so-called *self-noise* term. Notice that the watermark, defined as $\mathbf{w} \triangleq \mathbf{y} - \mathbf{x}$, is Λ-periodic both in $\mathbf{x}$ and $\mathbf{t}$ since it yields the same value for hosts and dither vectors of the form $\mathbf{x} + \mathbf{r},\ \mathbf{r} \in \Lambda$ and $\mathbf{t} + \mathbf{r},\ \mathbf{r} \in \Lambda$, respectively. The complete data hiding scheme is summarized in the block diagram of Fig. 1. The aim of the secret dither is just to apply a secret shift to the embedding lattice, and it does not change any of its fundamental properties concerning information transmission.[3] Nowadays, most of the lattice DC-DM schemes base their security on this strategy.

As it is usual in the analysis of quantization-based methods for data hiding [13], [16], a low embedding distortion regime is assumed, such that the variance of the host is much larger than the volume of the Voronoi region of Λ. The Voronoi region of a lattice Λ is denoted by $\mathcal{V}(\Lambda)$ and is defined as [19]

$$\mathcal{V}(\Lambda) \triangleq \{\mathbf{x} \in \mathbb{R}^n : Q_\Lambda(\mathbf{x}) = \mathbf{0}\}. \tag{4}$$

In practice, this assumption (which we will refer to in the sequel as the *flat-host assumption*) implies that the pdf of the host and that of the self-noise are approximately uniform inside each quantization cell and over $\mathcal{Z}(\Lambda) \triangleq (1-\alpha)\mathcal{V}(\Lambda)$, respectively. The flat-host assumption permits us to simplify the theoretical analysis, restricting our attention to the

[3] Strictly speaking, this is true only if the *flat-host assumption* (to be defined later) holds, as noted in [20].

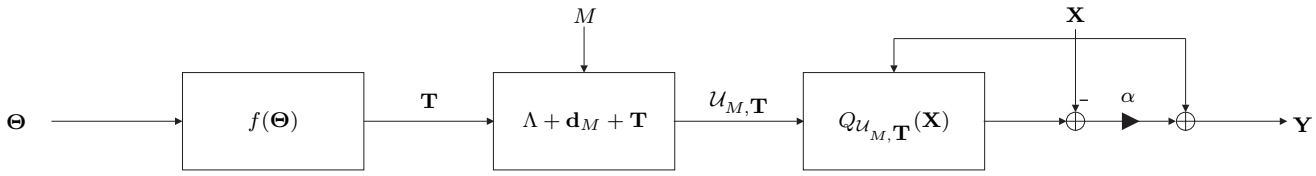


Fig. 1. Block diagram of lattice-based DC-DM with pseudo-random dithering. Parameter $\mathbf{\Theta}$ is the secret key.

modulo-reduced random variable $\tilde{\mathbf{Y}} \triangleq \mathbf{Y} \mod \Lambda = \mathbf{Y} - Q_\Lambda(\mathbf{Y})$.[4] Hence, the pdf of $\tilde{\mathbf{Y}}$ conditioned on the embedded message and the secret dither is

$$f(\tilde{\mathbf{y}}|m,\mathbf{t}) = \begin{cases} \text{vol}(\mathcal{Z}(\Lambda))^{-1}, & \tilde{\mathbf{y}} \in (\mathbf{d}_m + \mathbf{t} + \mathcal{Z}(\Lambda)) \bmod \Lambda \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

In our model, as is customary in theoretical analysis of watermarking methods, the host samples are considered independent and identically distributed (i.i.d.). Under these premises, a theoretical security analysis will be developed for the two scenarios (KMA, CMA) introduced in Section I. Obviously, the security level of the system depends on the statistical distribution of the secret dither, or better to say, of its modulo-Λ reduced version, $\tilde{\mathbf{T}}$. Due to the Λ-periodicity inherent in the watermark generation (see Eq. (3)), we have that $f(\tilde{\mathbf{y}}|\mathbf{T}=\mathbf{t}) = f(\tilde{\mathbf{y}}|\mathbf{T}=\mathbf{t}+\mathbf{r})\ \forall\ \mathbf{r} \in \Lambda$; hence

$$f(\tilde{\mathbf{y}}) = \int_{\mathbb{R}^n} f(\tilde{\mathbf{y}}|\mathbf{T}=\mathbf{t}) \cdot f(\mathbf{t}) d\mathbf{t} = \int_{\mathcal{V}(\Lambda)} f(\tilde{\mathbf{y}}|\mathbf{T}=\tilde{\mathbf{t}}) \cdot f(\tilde{\mathbf{t}}) d\tilde{\mathbf{t}}, \tag{6}$$

where $f(\tilde{\mathbf{t}}) = \sum_{\mathbf{r}\in\Lambda} f(\mathbf{t}+\mathbf{r})$ is the pdf of $\tilde{\mathbf{T}}$. This means that the pdf of the watermarked signal depends in last instance of the pdf of $\tilde{\mathbf{T}}$, and hence the secrecy of the codebook only depends on the statistics of $\tilde{\mathbf{T}}$.[5] Therefore, the support of $\mathbf{T}$ is bounded by $\mathcal{V}(\Lambda)$ hereinafter. We must note that $\mathbf{T}$ is usually assumed to be uniformly distributed over $\mathcal{V}(\Lambda)$ in most lattice data hiding schemes [13],[16], but this choice was not strictly motivated by security reasons, so it makes sense to wonder about its optimal distribution from this latter point of view.

A. Known Message Attack

When a sequence of watermarked signals $\{\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}\}$ and their associated messages $\{M_1, \ldots, M_{N_o}\}$ are observed, the information leakage about $\mathbf{T}$ can be calculated by means of the mutual information between the observations and the secret dither:

$$I(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}; \mathbf{T}|M_1, \ldots, M_{N_o}) = h(\mathbf{T}) - h(\mathbf{T}|\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}, M_1, \ldots, M_{N_o}), \tag{7}$$

where we have made use of the mutual independence between $\mathbf{T}$ and the embedded messages, also assumed to be mutually independent. Here, $h(\mathbf{T})$ is the differential entropy [6] of the random variable $\mathbf{T}$, and the second term of (7) is the *residual entropy* or *equivocation* of the dither after N_o observations, following Shannon's nomenclature [5]. The equivocation measures the remaining ignorance about the secret dither, so the appropriate distribution for

[4]It is worth noting that this modulo operation is virtually information-lossless [20, Sect. IV] in low embedding distortion regimes, as it is our case. This implies that the analysis is accurate, in the sense that $I(\tilde{\mathbf{Y}};\mathbf{T}) \approx I(\mathbf{Y};\mathbf{T})$.

[5]However, attacks at a cryptographic level would be indeed interested in knowing the exact value of $\mathbf{t}$.

$\mathbf{T}$ should be chosen in order to maximize this value. To this end, let us consider the conditional pdf of the dither: the statistical properties of the watermarked signals give rise to the notion of *feasible region* of the dither, formally defined as the support of its conditional pdf after N_o observations. The next property will be widely used throughout the text.

Property 1: Boundedness of the feasible region. The feasible region is bounded by $\mathcal{S}_{N_o} \triangleq \bigcap_{i=1}^{N_o} \mathcal{D}_i$, where

$$\mathcal{D}_i \triangleq (\tilde{\mathbf{y}}_i - \mathbf{d}_{m_i} - \mathcal{Z}(\Lambda)) \mod \Lambda,\ i = 1, \ldots, N_o, \tag{8}$$

Proof: Application of Bayes' rule yields

$$f(\mathbf{t}|\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o}, m_1, \ldots, m_{N_o}) = \frac{f(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o}, m_1, \ldots, m_{N_o}|\mathbf{t}) \cdot f(\mathbf{t})}{f(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o}, m_1, \ldots, m_{N_o})} = \frac{f(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o}|m_1, \ldots, m_{N_o}, \mathbf{t}) \cdot f(\mathbf{t})}{f(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o}|m_1, \ldots, m_{N_o})}, \tag{9}$$

where $\tilde{\mathbf{y}}_i \in (\mathbf{d}_m + \mathbf{t} + \mathcal{Z}(\Lambda)) \mod \Lambda,\ i = 1, \ldots, N_o$. Notice that each random variable $\tilde{\mathbf{Y}}_i$ is a function of the triple $(\mathbf{X}_i, M_i, \mathbf{T})$, and the host samples $\mathbf{X}_i$ in our model are mutually independent. This means that the observations $\{\tilde{\mathbf{Y}}_i\}$ are conditionally independent given the dither; hence, Eq. (9) can be rewritten as

$$f(\mathbf{t}|\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o}, m_1, \ldots, m_{N_o}) = \frac{f(\mathbf{t}) \cdot \prod_{i=1}^{N_o} f(\tilde{\mathbf{y}}_i|m_i, \mathbf{t})}{f(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o}|m_1, \ldots, m_{N_o})} \tag{10}$$

$$= \frac{f(\mathbf{t}) \cdot \prod_{i=1}^{N_o} f((\tilde{\mathbf{y}}_i - \mathbf{d}_{m_i} - \mathbf{t}) \mod \Lambda|M_i = 0, \mathbf{T} = \mathbf{0})}{f(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o}|m_1, \ldots, m_{N_o})}, \tag{11}$$

where (11) follows from the flat-host assumption. By recalling Eq. (5), it is clear that each term in the numerator of (11) is nonzero iff $(\tilde{\mathbf{y}}_i - \mathbf{d}_{m_i} - \mathbf{t}) \mod \Lambda \in \mathcal{Z}(\Lambda)$, or equivalently, iff $\mathbf{t} \in \mathcal{D}_i$, with $\mathcal{D}_i$ given by (8). Hence, it is clear that the feasible region of $\mathbf{t}$ is contained in $\bigcap_{i=1}^{N_o} \mathcal{D}_i$, independently of the distribution of $\mathbf{T}$. ■

Property 1 allows us to state the following lemma.

Lemma 1: Maximization of the residual entropy. The residual entropy is maximized for $\mathbf{T} \sim U(\mathcal{V}(\Lambda))$, yielding a conditional pdf uniformly distributed in $\mathcal{S}_{N_o}$, that is

$$f(\mathbf{t}|\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o}, m_1, \ldots, m_{N_o}) = \begin{cases} (\mathrm{vol}\,(\mathcal{S}_{N_o}))^{-1}, & \mathbf{t} \in \mathcal{S}_{N_o} \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Proof: By the definition of residual entropy, we have

$$h(\mathbf{T}|\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}, M_1, \ldots, M_{N_o}) = E[h(\mathbf{T}|\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o}, m_1, \ldots, m_{N_o})], \tag{13}$$

where the expectation is taken over the joint pdf $f(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o}, m_1, \ldots, m_{N_o})$. Since the feasible region of the dither is bounded by $\mathcal{S}_{N_o}$, its entropy will be maximized when the dither is uniformly distributed in $\mathcal{S}_{N_o}$, i.e,

$$h(\mathbf{T}|\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}, M_1, \ldots, M_{N_o}) = -E[\log(\mathbf{T}|\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o}, m_1, \ldots, m_{N_o})] \leq E[\log(\mathrm{vol}(\mathcal{S}_{N_o}))]. \tag{14}$$

Since the denominator of (11) does not depend on $\mathbf{t}$, then the choice $\mathbf{T} \sim U(\mathcal{V}(\Lambda))$ suffices for achieving such distribution, and hence equality in (14). ■

The optimal distribution resulting from Lemma 1 also brings additional desirable properties: it provides statistical independence between the self-noise and the host signal [21], and most importantly, it does not prevent from achieving

capacity in the Gaussian channel in the asymptotic set-up ($n \to \infty$) [22]. Hence, the choice of $\mathbf{T} \sim U(\mathcal{V}(\Lambda))$ is *good* from the robustness and security points of view, and this will be the chosen distribution in the remaining of this paper unless otherwise stated. Hence, by combining Property 1 and Lemma 1, the residual entropy results in

$$h(\mathbf{T}|\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}, M_1, \ldots, M_{N_o}) = E[\log(\text{vol}(\mathcal{S}_{N_o}))], \tag{15}$$

where the expectation is taken over the joint pdf of the observations. In case of one observation ($N_o = 1$) we have

$$h(\mathbf{T}|\tilde{\mathbf{Y}}_1, M_1) = \log(\text{vol}(\mathcal{Z}(\Lambda))) = \log((1-\alpha)^n \text{vol}(\mathcal{V}(\Lambda))), \tag{16}$$

and the information leakage is given by

$$I(\tilde{\mathbf{Y}}_1; \mathbf{T}|M_1) = h(\mathbf{T}) - h(\mathbf{T}|\tilde{\mathbf{Y}}_1, M_1) = -n\log(1-\alpha) \tag{17}$$

for all $\alpha \in [0, 1]$, independently of the specific lattice chosen for embedding. This result clearly shows a trade-off between security and achievable rate: theoretical analyses [16], [22] show that, in AWGN channels, the value of α must approach 1 for maximizing the achievable rate in the high-WNR region; however, bear in mind that for $\alpha \approx 1$, one observation suffices to get an accurate estimate of the centroids in $\mathcal{U}_m$, and consequently of the secret dither, due to the structure imposed to the codebook. This is reflected in the residual entropy of the dither (16), for which $\lim_{\alpha \to 1} h(\mathbf{T}|\tilde{\mathbf{Y}}_1, M_1) = -\infty$.

For $N_o > 1$, one must consider two different cases: 1) for $\alpha = 1$, the mutual information is maximum for $N_o = 1$, as we have just discussed, so more observations will not provide additional information about $\mathbf{T}$ (i.e., it becomes deterministic for $N_o = 1$); 2) for $\alpha < 1$, the mutual information does not increase linearly due to the dependence between observations. Its general behavior is stated in the following Lemma.

Lemma 2: If $\alpha < 1$, the mutual information about the secret dither is an increasing, concave function of the number of observations N_o.

Proof: To see that the mutual information is always increasing, consider the function

$$\Delta I(N) = I(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N+1}; \mathbf{T}|M_1, \ldots, M_{N+1}) - I(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_N; \mathbf{T}|M_1, \ldots, M_N), \tag{18}$$

which is nothing but the average information about $\mathbf{t}$ that is gained with the $(N+1)$-th observation. Such function is easily seen to be always non-negative:

$$\Delta I(N) = h(\mathbf{T}|\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_N, M_1, \ldots, M_N) - h(\mathbf{T}|\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N+1}, M_1, \ldots, M_{N+1}) \geq 0, \tag{19}$$

where (19) follows from the fact that conditioning reduces entropy [6]. In this case strict inequality holds in (19), due to (12) and (13) (i.e., the mean volume of $\mathcal{S}_{N_o}$ is always reduced with each new observation, with the obvious exception of deterministic $\mathbf{t}$). Thus, the mutual information is always increasing.

In order to prove the concavity of the mutual information, we make use of the following claim [23]

Claim: Discrete concavity. A discrete function $f(k)$, with $k \in \mathbb{Z}$, is (strictly) concave if and only if $\Delta f(k)$ is (decreasing) non-increasing, with $\Delta f(k) = f(k+1) - f(k)$.

Thus, the mutual information will be (strictly) concave iff $\Delta I(N)$ is (decreasing) non-increasing. By the definition of mutual information, we have

$$I(\tilde{\mathbf{Y}}_1,\ldots,\tilde{\mathbf{Y}}_N;\mathbf{T}|M_1,\ldots,M_N) = h(\tilde{\mathbf{Y}}_1,\ldots,\tilde{\mathbf{Y}}_N|M_1,\ldots,M_N) - h(\tilde{\mathbf{Y}}_1,\ldots,\tilde{\mathbf{Y}}_N|M_1,\ldots,M_N,\mathbf{T}). \tag{20}$$

Making use of the chain rule for entropies [6, Section 9.6] in (20), we can write

$$\Delta I(N+1) - \Delta I(N) = h(\tilde{\mathbf{Y}}_{N+2}|\tilde{\mathbf{Y}}_1,\ldots,\tilde{\mathbf{Y}}_{N+1},M_1,\ldots,M_{N+2}) - h(\tilde{\mathbf{Y}}_{N+1}|\tilde{\mathbf{Y}}_1,\ldots,\tilde{\mathbf{Y}}_N,M_1,\ldots,M_{N+1}) \leq 0, \tag{21}$$

taking into account again that conditioning reduces entropy. This concludes the proof of the lemma. ■

B. Constant Message Attack

The easiest way of addressing this scenario is to regard it as a collection of several KMA problems. When the message embedded is unknown but unchanged for the whole sequence of observations, the conditional pdf of the dither after N_o observations can be expressed as

$$f(\mathbf{t}|\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{N_o},\text{CM}) = \frac{1}{|\mathcal{M}|}\sum_{m=0}^{|\mathcal{M}|-1} f(\mathbf{t}|\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{N_o},m,\ldots,m), \tag{22}$$

where CM stands for *constant message*, and $|\mathcal{M}|$ denotes the size of the alphabet. This means that the feasible region $\mathcal{S}_{N_o}^{CMA}$ for the dither in the CMA case is simply the union of the feasible regions of $|\mathcal{M}|$ KMA problems. Formally,

$$\mathcal{S}_{N_o}^{CMA} = \bigcup_{m=0}^{|\mathcal{M}|-1} (\mathcal{S}_{N_o} + \mathbf{d}_m), \tag{23}$$

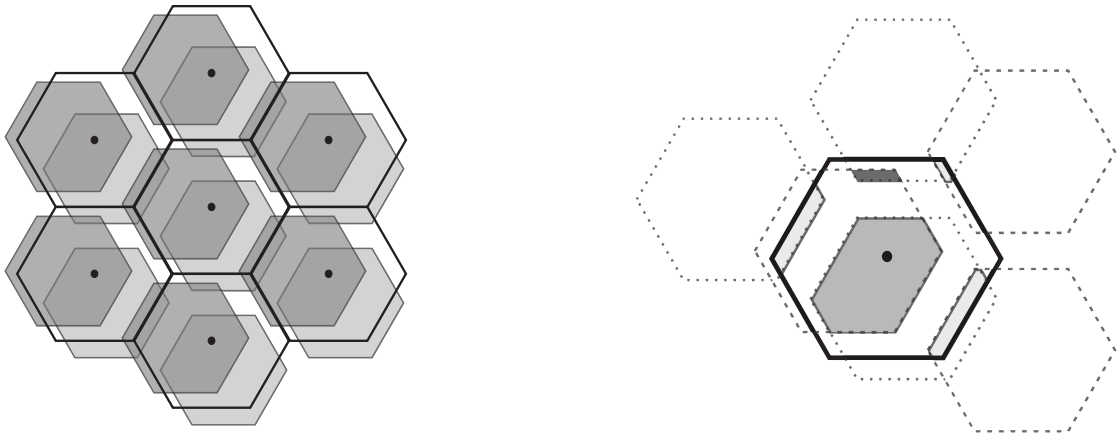
with $\mathcal{S}_{N_o}$ defined in Property 1. Using the Borel-Cantelli Lemma, and under the assumptions stated in this paper, it can be shown that $\text{vol}(\mathcal{S}_{N_o})$ converges to zero *almost surely* when $N_o \to \infty$; thus, the different regions that constitute $\mathcal{S}_{N_o}^{CMA}$ will be disjoint for sufficiently large N_o; in such case, the residual entropy is again maximized if $\mathbf{T} \sim U(\mathcal{V}(\Lambda))$ is chosen, but it is not necessarily the optimal distribution for all N_o. Due to (22), the residual entropy can be upper bounded as

$$h(\mathbf{T}|\tilde{\mathbf{Y}}_1,\ldots,\tilde{\mathbf{Y}}_{N_o},\text{CM}) \leq h(\mathbf{T}|\tilde{\mathbf{Y}}_1,\ldots,\tilde{\mathbf{Y}}_{N_o},M_1,\ldots,M_{N_o}) + \log(|\mathcal{M}|), \tag{24}$$

resulting in a lower bound to the information leakage. Equality in (24) is achieved when the regions $\mathcal{S}_{N_o} + \mathbf{d}_m$ are disjoint, which means that, as N_o increases, the bound will be asymptotically tight. However, if the value of α is above a certain threshold (which depends on the lattice partition) such regions are always disjoint, and the bound is reached for all N_o; this is the case, for instance, when $\alpha > \alpha_T = 1 - \frac{1}{|\mathcal{M}|}$, for self-similar partitions [7], [22].

III. LATTICE COMPARISON

This section tries to shed some light on two fundamental questions: 1) given n, what is the *best* lattice (if any) in terms of security; and 2) does an increase of n improve the security level. The discussion will be focused on the KMA problem, although it can be extended to the CMA scenario by taking into account the remarks made in Section II-B. Before proceeding with the analysis, we need to introduce the following definition:



(a) (b)

Fig. 2. Illustration of a non-connected feasible region for two observations using small α. In (a), the solid lines are the Voronoi regions of Λ, and the feasible regions for the centroids defined by each observation are the shaded ones. The figure depicted in (b) is the modulo-Λ reduction of the intersection between the shaded regions in (a), showing three resulting modulo-Λ convex regions (illustrated with different shadings).

Definition 1: A set $\mathcal{S}$ is said to be modulo-Λ convex if there exists $\mathbf{r}$ such that $\mathcal{S} - Q_{\Lambda+\mathbf{r}}(\mathcal{S})$ is convex.

The notion of modulo-Λ convexity is key to our analysis, due to the next property.

Property 2: For $\alpha \geq 0.5$, the feasible region $\mathcal{S}_{N_o}$ is always a modulo-Λ convex set.

Proof: Let us define

$$\tilde{\mathbf{V}}_i \triangleq (\tilde{\mathbf{Y}}_i - \mathbf{D}_{m_i} - \mathbf{T}) \mod \Lambda \tag{25}$$

and $\mathcal{D}'_i \triangleq \tilde{\mathbf{V}}_i - \mathcal{Z}(\Lambda),\ i = 1, \ldots, N_o$. By recalling the generation of the watermarked signal (3), it is clear that $\tilde{\mathbf{V}}_i \sim U(\mathcal{Z}(\Lambda))$. If $\alpha \geq 0.5$, then $\tilde{\mathbf{V}}_i + \mathbf{r} \in \mathcal{V}(\Lambda),\ \forall\ \mathbf{r} \in \mathcal{Z}(\Lambda)$. Hence, $\mathcal{D}'_i \subset \mathcal{V}(\Lambda)$, and obviously $\bigcap_i \mathcal{D}'_i \subset \mathcal{V}(\Lambda)$. Since $\mathcal{D}'_i,\ i = 1, \ldots, N_o$, are convex sets, their intersection is also a convex set. Taking into account that $\mathcal{D}_i = (\mathbf{T} + \mathcal{D}'_i) \bmod \Lambda$, the property follows. ■

The use of $\alpha < 0.5$ may lead to non-convex feasible regions, as illustrated in Fig. 2-(b), where the feasible region for the dither is composed of three different modulo-Λ sets. However, as can be seen in the proof of Property 2, under the assumption of $\alpha \geq 0.5$ it is possible to find a shifted version of the problem such that the feasible region is always modulo-Λ convex, according to Definition 1. This property permits us to drop out the modulo operation from the expressions of the feasible regions. Bear in mind that the entropy is invariant to translations, so this simplification does not change the results. In order to provide a fair comparison between different lattices, they are scaled so as to present the same embedding distortion, which due to the flat-host assumption is given by

$$D_w = \frac{\alpha^2}{n} E\{\mathbf{q}^T \mathbf{q}\} = \frac{\alpha^2 \int_{\mathcal{V}(\Lambda)} ||\mathbf{q}||^2 d\mathbf{q}}{n \cdot \text{vol}(\mathcal{V}(\Lambda))} = \alpha^2 P(\Lambda), \tag{26}$$

where $\mathbf{q} = Q_\Lambda(\mathbf{x}) - \mathbf{x}$ is the quantization error, and $P(\Lambda)$ is the second order moment per dimension of $\mathcal{V}(\Lambda)$. For computing the residual entropy, the expectation in (15) must be taken over $f(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o}, m_1, \ldots, m_{N_o})$, but the conditional pdf of $\mathbf{T}$, given by (12), does not depend on the specific sequence of messages embedded, as long as the latter is known; this implies that, for the expectations, the message sequence can be assumed to be deterministic. Since it is not always possible to obtain closed-form expressions for the information leakage (even for low-dimensional lattices), we must resort in general to Monte Carlo integration and bounding techniques.

A. Exact computation for the cubic lattice

For the scaled cubic lattice[6] $\Delta\mathbb{Z}^n = (x_1, \ldots, x_n),\ x_i \in \Delta\mathbb{Z}$ it is possible to obtain a closed-form expression for the residual entropy. From Eq. (15), the residual entropy is given by the expectation of the log-volume of the feasible region for the dither. Since the latter for the cubic lattice is always a hyperrectangle, using Property 2 we can write

$$E[\log(\text{vol}(\mathcal{S}_{N_o}))] = \sum_{k=1}^{n} E[\log(W_k)] = n \cdot E[\log(W)], \tag{27}$$

where W_k is the random variable that measures the length of the feasible interval in the k-th dimension, and the last equality follows because the quantization step is the same for all dimensions. The random variable W is given by

$$W = \text{vol}(\bigcap_{i=1}^{N_o} (\tilde{V}_i - \mathcal{I})), \tag{28}$$

with $\tilde{V}_i$ a random variable uniformly distributed in $\mathcal{I} \triangleq [-(1-\alpha)\Delta/2, (1-\alpha)\Delta/2)$. Hence, the problem is reduced to a scalar subproblem consisting in computing $E[\log(W)]$, i.e., the residual entropy in one dimension. This result is used in Appendix I, under the assumption of $\alpha \geq 0.5$, to show that the residual entropy per dimension is given by

$$\frac{1}{n} h(\mathbf{T}|\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}, M_1, \ldots, M_{N_o}) = \log((1-\alpha)\Delta) - H_{N_o} + 1 = \log(\sqrt{12} D_w) - H_{N_o} + 1 + \log\left(\frac{1-\alpha}{\alpha}\right), \tag{29}$$

where $H_{N_o} \triangleq \sum_{i=1}^{N_o} \frac{1}{i}$ is the N_o-th harmonic number, D_w is the embedding distortion according to (26), and we have taken into account that for the cubic lattice $P(\Lambda) = \Delta^2/12$.

B. Monte Carlo integration

When the analytical evaluation of (15) becomes intractable we resort to Monte Carlo integration. The fact that the feasible region is reduced with each new observation makes necessary an additional task of computing a tight region of integration so as to preserve the accuracy of the Monte Carlo method (as will be seen in step 3 of the algorithm outlined below). In order to give a comparison between different standard lattices, we consider the root lattices and their duals (the best known lattice quantizers for $n \leq 8$), namely A_2 (hexagonal lattice), D_3, $D_4 \cong D_4^*$, D_5, E_7, $E_8 \cong E_8^*$. For their definition and properties, see [19], [24]. All these lattices are scaled so as to present the same embedding distortion per dimension as the cubic lattice $\Delta\mathbb{Z}^n$ with $\Delta = 1$, that is, $1/12$.

The procedure followed for the Monte Carlo simulations is briefly outlined here.

1) We assume without loss of generality that $\mathbf{t} = \mathbf{0}$. Hence, a sequence of N_o observed vectors uniformly distributed in $(1-\alpha)\Delta\mathcal{V}(\Lambda)$, with Δ such that $P(\Lambda) = 1/12$, is generated.

2) $\mathcal{V}(\Lambda)$ is outer bounded by a hypercube whose edge length is twice the covering radius [19] of Λ. This gives an outer bound to $\mathcal{D}_i$ (Eq. (8)), which is used to compute an outer approximation $\mathcal{S}_{N_o}^u$ of the feasible region.

3) The feasible region resulting from the previous step (which is a hyperrectangle) is shrunk along each dimension so as to tightly bound the true feasible region $\mathcal{S}_{N_o}$. This is accomplished by means of a bisection algorithm which looks for the tightest limits of the outer bounding hyperrectangle in each dimension. The need for this step is justified

[6]We consider the same quantization step in each dimension, although the results can be straightforwardly extended to a general case.

by the fact that, for large N_o, the ratio vol$(\mathcal{S}^u_{N_o})$/vol$(\mathcal{S}_{N_o})$ becomes too large, affecting the accuracy of Monte Carlo integration.

4) A large number of points uniformly distributed in the hyperrectangle of the previous step is generated. For each of these points, it is checked whether it belongs to $\bigcap_{i=1}^{N_o} \mathcal{D}_i$; if so, the considered point belongs to $\mathcal{S}_{N_o}$. Finally, the log-volume of $\mathcal{S}_{N_o}$ is computed by Monte Carlo integration, and the residual entropy is obtained by averaging the log-volume over a large number of realizations. In steps 3) and 4), fast quantizing algorithms [25] are used.

The results of Monte Carlo integration indicate that the lattice Λ_n^* that maximizes the residual entropy for each n is that with the best mean-squared quantization properties. This can be formally expressed as

$$\begin{aligned} \Lambda_n^* = \quad & \arg\min_{\Lambda \in \mathcal{L}_n} G(\Lambda) \\ & \text{subject to } P(\Lambda) = constant \end{aligned} \tag{30}$$

where $\mathcal{L}_n$ is the set of root lattices of dimensionality $n \leq 8$, and $G(\Lambda) \triangleq \frac{P(\Lambda)}{\text{vol}(\mathcal{V}(\Lambda))^{2/n}}$ is the normalized second order moment of Λ. Notice that Λ_n^* maximizes vol$(\mathcal{V}(\Lambda))$ for given n and $P(\Lambda)$, and consequently Λ_n^* has the highest a priori entropy in $\mathcal{L}_n$, due to the uniformity of $\mathbf{T}$. For illustration purposes, Fig. 3 gives a comparison between the residual entropy per dimension using the cubic lattice and that using some of the root lattices. Although we do not claim that the above result holds for the whole set of lattices with arbitrary n, at least it suggests that the security level of a lattice data hiding scheme can be improved by increasing n and choosing the lattice Λ with the lowest $G(\Lambda)$. This leads us to conjecture that a hypothetical spherically-shaped Voronoi region will provide an upper bound to the residual entropy, since the sphere is the region of $\mathbb{R}^n$ with the smallest normalized second order moment. This is indeed so for the set of lattices considered in our experiments: as an example, the result obtained with the 8-dimensional sphere (also obtained through Monte Carlo) is plotted in Fig. 3. Unfortunately, the space can not be tessellated with spherical regions (except for $n = 1$), so it is not possible to construct *spherical* lattice quantizers; nevertheless, as it was shown in [26], as n increases there exist lattices whose normalized second order moment tend to that of a sphere.[7] The security of lattice DC-DM using this type of lattices is studied in the next section.

C. Bounds and asymptotics on the equivocation for "good" lattices

Throughout this section, we will make use of two assumptions: 1) $\alpha \geq 0.5$; 2) we are using Λ_n^*, the optimal (in a mean-squared error sense) lattice quantizer in n-dimensions. As discussed in the proof of Property 2, Assumption 1 makes the modulo operation transparent for the computation of the entropy, since this is invariant to translations. Making use of the chain rule for mutual informations [6] we can write

$$\begin{aligned} I(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}; \mathbf{T} | M_1, \ldots, M_{N_o}) &= I(\tilde{\mathbf{Y}}_1; \mathbf{T} | M_1) + I(\tilde{\mathbf{Y}}_2, \ldots, \tilde{\mathbf{Y}}_{N_o}; \mathbf{T} | \tilde{\mathbf{Y}}_1, M_1, \ldots, M_{N_o}) \\ &= I(\tilde{\mathbf{Y}}_1; \mathbf{T} | M_1) + I(\tilde{\mathbf{Y}}_2, \ldots, \tilde{\mathbf{Y}}_{N_o}; \mathbf{T}' | M_2, \ldots, M_{N_o}), \end{aligned} \tag{31}$$

[7] Moreover, this is a necessary condition for the lattices in order to achieve the channel capacity in the lattice DC-DM scheme [22].

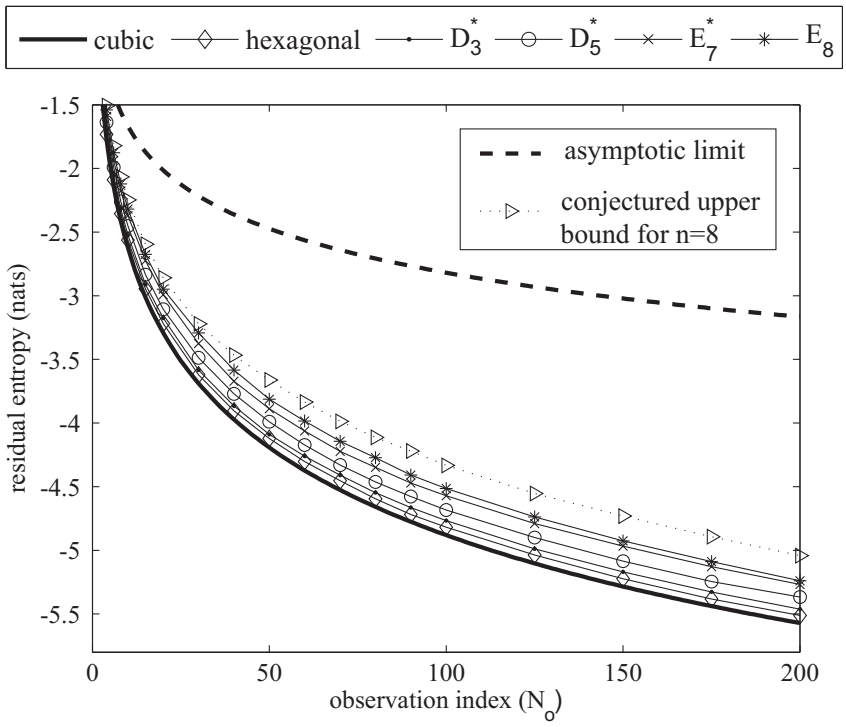


Fig. 3. Residual entropies per dimension for different lattices. All plots for the root lattices (but for the cubic one, which is theoretical) were obtained through Monte Carlo integration. The asymptotic limit corresponds to Eq. (35). The embedding distortion in all cases is $D_w = \alpha^2/12$, with $\alpha = 0.5$.

where $\mathbf{T}' \sim U((1-\alpha)\mathcal{V}(\Lambda_n^*))$ is the dither conditioned on the first observation (as it follows from Property 1 and Lemma 1). Thus, each new observation conditioned on $\tilde{\mathbf{Y}}_1$ and M_1 can be written as[8]

$$\tilde{\mathbf{Y}}_i = \mathbf{Z}_i + \mathbf{T}' + \mathbf{d}_{m_i},\ i = 2, \ldots, N_o, \tag{32}$$

where $\mathbf{Z}_i \triangleq (1-\alpha)(\mathbf{X}_i - Q_{\Lambda_n^*}(\mathbf{X}_i))$ is the self-noise term, with the same statistical distribution as $\mathbf{T}'$, and hence with second moment per dimension $(1-\alpha)^2 P(\Lambda_n^*)$. From Eq. (31), it can be seen that the following equality holds:

$$h(\mathbf{T}|\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}, M_1, \ldots, M_{N_o}) = h(\mathbf{T}'|\tilde{\mathbf{Y}}_2, \ldots, \tilde{\mathbf{Y}}_{N_o}, M_2, \ldots, M_{N_o}),\ \text{for } N_o \geq 2, \tag{33}$$

so we can use the second term of (31) for obtaining a lower bound on the equivocation per dimension, as shown in Appendix II:

$$\frac{1}{n} h(\mathbf{T}'|\tilde{\mathbf{Y}}_2, \ldots, \tilde{\mathbf{Y}}_{N_o}, M_2, \ldots, M_{N_o}) \geq \frac{N_o}{2} \log\left(\frac{P(\Lambda_n^*)}{G(\Lambda_n^*)}\right) - \frac{(N_o-1)}{2} \log\left(2\pi e P(\Lambda_n^*)\right) - \frac{1}{2}\log(N_o) + \log(1-\alpha). \tag{34}$$

This lower bound is loose for small n, but the next result shows that it is asymptotically tight for $n \to \infty$.

Theorem 1: In the limit when $n \to \infty$, using the optimum lattice quantizer Λ_n^*, the equivocation per dimension in lattice DC-DM is given by

$$\lim_{n\to\infty} \frac{1}{n} h(\mathbf{T}'|\tilde{\mathbf{Y}}_2, \ldots, \tilde{\mathbf{Y}}_{N_o}, M_2, \ldots, M_{N_o}) = \frac{1}{2}\log(2\pi e D_w) - \frac{1}{2}\log(N_o) + \log\left(\frac{1-\alpha}{\alpha}\right),\ \text{for } N_o \geq 2, \tag{35}$$

where D_w is the embedding distortion per dimension (26).

[8]As discussed before, the residual entropy in the KMA scenario does not depend on the specific message sequence as long as this is known, so we consider $\mathbf{d}_{m_i} = \mathbf{0}\ \forall\ i = 1, \ldots, N_o$, without loss of generality for the remaining of this section and in the corresponding appendices.

Proof: See Appendix III. ■

Notice that when $n \to \infty$, (34) coincides with (35), because $G(\Lambda_n^*) \to 1/2\pi e$. The first term in (35) accounts for the relation between the embedding distortion and the a priori entropy of the secret dither. The second term tells us how the equivocation decreases with N_o, and the third term shows the dependence with the distortion compensation parameter α, which basically introduces a constant shift in the equivocation curve (recall that for $\alpha = 1$, the residual entropy is $-\infty$ for $N_o \geq 1$). The asymptotic value of the equivocation is plotted in Figure 3 for reference, showing the gap with the root lattices studied before. The above theorem is the formal statement of a more intuitive result: the Voronoi region of Λ_n^* tends to a sphere, and in turn the uniform distribution in $\mathcal{V}(\Lambda_n^*)$ tends asymptotically to a Gaussian distribution (in the normalized entropy sense) [26]; hence, roughly speaking, each modulo-Λ reduced observation (Eq. (32)) becomes closer to a Gaussian distribution with variance D_w/α^2, whose mean is given by the secret dither (also with the same statistical distribution). This interpretation brings more insight in the comparison of the theoretical security between lattice DC-DM and additive spread spectrum methods. For the latter, the embedding function is given by $\mathbf{Y} = \mathbf{X} + (-1)^m \mathbf{U}$, where $\mathbf{X}$ and $\mathbf{U}$ are the host and the spreading vector, respectively, with the latter playing the role of the secret key. Notice that the resemblance between this embedding function and (32) implies similar security properties for both methods. Considering that $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \sigma_X^2 \cdot \mathbf{I}_n)$ and $\mathbf{U} \sim \mathcal{N}(\mathbf{0}, \sigma_U^2 \cdot \mathbf{I}_n)$, it was shown in [3] that

$$\frac{1}{n} h(\mathbf{U}|\mathbf{Y}_1, \ldots, \mathbf{Y}_{N_o}, M_1, \ldots, M_{N_o}) = \frac{1}{2} \log\left(2\pi e D_w\right) - \frac{1}{2} \log\left(1 + N_o \frac{D_w}{\sigma_X^2}\right), \tag{36}$$

where now $D_w = \sigma_U^2$. It can be readily seen that the decrease in the equivocation for additive spread spectrum is determined by the ratio σ_U^2/σ_X^2, which is usually very small due to imperceptibility constraints. Instead, for lattice DC-DM after the modulo-Λ reduction, the power of both the watermark and the host interference are the same, i.e., $(1-\alpha)^2 P(\Lambda_n^*)$; this explains the term $\frac{1}{2}\log(N_o)$ in (35) and the rapid decrease of the equivocation, compared to that of (36).

Fig. 4 shows a comparison between lattice DC-DM and additive spread spectrum for different values of embedding distortion, parameterized by the Document to Watermark Ratio, defined as DWR $\triangleq 10 \log_{10}(\sigma_X^2/D_w)$.

D. Bounds on the estimation error

Let us define the estimation error as $\mathbf{e} \triangleq \mathbf{t} - \hat{\mathbf{t}}$, where $\hat{\mathbf{t}}$ is the dither estimate. If the covariance matrix of the estimation error is given by $\mathbf{R}_E$, then it is immediate to upper bound its entropy by

$$h(\mathbf{E}) \leq \frac{1}{2} \log\left((2\pi e)^n |\mathbf{R}_E|\right). \tag{37}$$

Furthermore, note that

$$h(\mathbf{E}) = h(\mathbf{T} - \hat{\mathbf{T}}) \geq h(\mathbf{T} - \hat{\mathbf{T}}|\mathbf{Y}_1, \ldots, \mathbf{Y}_{N_o}, M_1, \ldots, M_{N_o}) = h(\mathbf{T}|\mathbf{Y}_1, \ldots, \mathbf{Y}_{N_o}, M_1, \ldots, M_{N_o}), \tag{38}$$

since $\hat{\mathbf{T}}$ is a function of the observations. Thus,

$$h(\mathbf{T}|\mathbf{Y}_1, \ldots, \mathbf{Y}_{N_o}, M_1, \ldots, M_{N_o}) \leq \frac{1}{2} \log\left((2\pi e)^n |\mathbf{R}_E|\right) \leq \frac{n}{2} \log\left(2\pi e \frac{\mathrm{tr}(\mathbf{R}_E)}{n}\right), \tag{39}$$

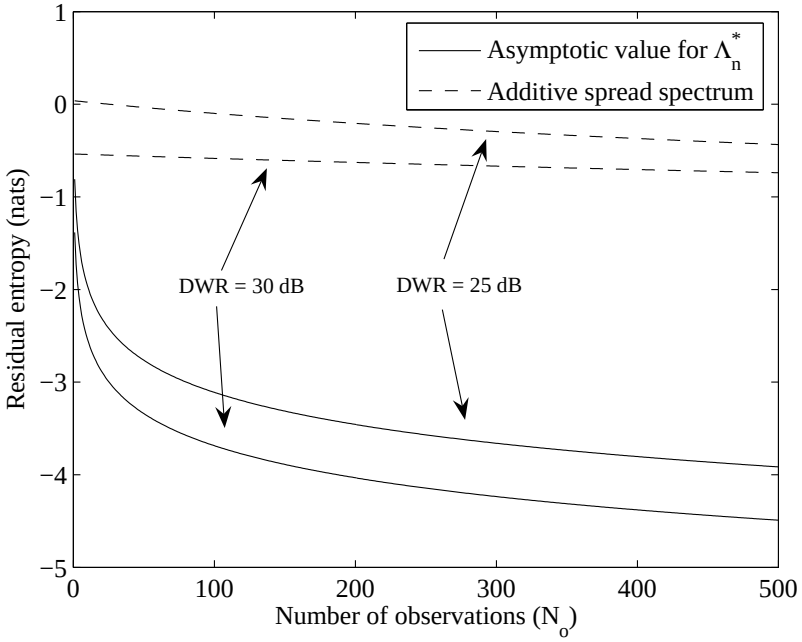


Fig. 4. Comparison, in terms of equivocation per dimension, between lattice DC-DM and additive spread spectrum. $\alpha = 0.7$ for DC-DM.

where the second inequality follows from the fact that $|\mathbf{R}_E|^{\frac{1}{n}} \leq \frac{\mathrm{tr}(\mathbf{R}_E)}{n}$ [6, Th. 16.8.4]. Let us define the variance per dimension of the estimation error as $\sigma_E^2 \triangleq \frac{\mathrm{tr}(\mathbf{R}_E)}{n}$. Then, from (39) we have the following lower bound on σ_E^2:

$$\sigma_E^2 \geq \frac{1}{2\pi e} e^{\frac{2}{n} h(\mathbf{T}|\mathbf{Y}_1,\ldots,\mathbf{Y}_{N_o},M_1,\ldots,M_{N_o})}, \tag{40}$$

which is nothing but the entropy power of $\mathbf{T}$ given N_o observations [6]. It can be observed that, for achieving an error-free estimate, the equivocation must necessarily approach $-\infty$. Substituting Eq. (35) into (40), we arrive at the following bound for $n \to \infty$ and the optimal lattice quantizer:

$$\sigma_E^2 \geq \frac{(1-\alpha)^2 P(\Lambda_n^*)}{N_o}, \tag{41}$$

The above bound is attained using the simple averaging estimator, but taking into account that the observations must be properly shifted in order to avoid problems with the modulo-Λ reduction; thus, if we define

$$\tilde{\mathbf{v}}_i = (\tilde{\mathbf{y}}_i - \mathbf{d}_{m_i} - \tilde{\mathbf{y}}_1 + \mathbf{d}_{m_1}) \bmod \Lambda,\ i = 1, \ldots, N_o, \tag{42}$$

then the optimal dither estimator for $\Lambda_n^*, n \to \infty$, is given by

$$\hat{\mathbf{t}}_{av} = \left(\tilde{\mathbf{y}}_1 - \mathbf{d}_{m_1} + \frac{1}{N_o} \sum_{i=1}^{N_o} \tilde{\mathbf{v}}_i \right) \bmod \Lambda. \tag{43}$$

The achievability of (41) follows from the fact that, for Λ_n^*, the self-noise and the secret dither follow asymptotically a Gaussian distribution as $n \to \infty$. Thus, this result about the estimation error can be compared to the estimation error for the cubic lattice; since we are interested in computing the behavior for large N_o, we make use of the approximation $H_{N_o} \approx \log(N_o) + \gamma$, which is asymptotically tight for large N_o, with $H_{N_o} \triangleq \sum_{i=1}^{N_o} \frac{1}{i}$ the harmonic number and γ the Euler-Mascheroni constant, defined as $\gamma \triangleq \lim_{N_o \to \infty} H_{N_o} - \log(N_o)$. In this case we have, using (29)

$$\sigma_E^2 \geq \frac{1}{2\pi e} e^{2(\log((1-\alpha)\Delta) - H_{N_o} + 1)} \approx \frac{1}{2\pi e} e^{2(\log((1-\alpha)\Delta) - \log(N_o) + 1 - \gamma)} = \frac{1}{2\pi e^{2\gamma - 1}} \cdot \frac{(1-\alpha)^2 \Delta^2}{N_o^2}. \tag{44}$$

Thus, the variance per dimension approximately decreases with the inverse of the squared number of observations. This bound can even be compared to the exact error variance of the optimal dither estimator, in order to check the tightness of the bound. For the cubic lattice, dither estimation may be carried out independently for each component without loss of optimality. It is a well known result that the optimal dither estimator in a mean-squared error sense is given by the mean value of the dither conditioned on the N_o observations: in our case, the i-th component of the dither is uniformly distributed in an interval $[x_1, x_2]$; hence, the optimal estimate is $\hat{t} = (x_1 + x_2)/2$, and the variance per dimension of the estimation error is

$$\sigma_E^2 = E\left[(T - \hat{t})^2\right] = \text{var}(T) = \frac{1}{12} \cdot E[W^2], \tag{45}$$

where $w = |x_2 - x_1|$ is the width of the feasible interval, and the expectation is taken over the joint pdf of the observations. Actually, this expectation may be computed by replacing $\log(w)$ by w^2 in Eq. (64) of Appendix I, resulting in

$$\sigma_E^2 = \frac{1}{2} \cdot \frac{(1-\alpha)^2 \Delta^2}{2 + 3N_o + N_o^2}, \tag{46}$$

which for large N_o is dominated by the term N_o^2, differing from the right hand side of (44) only in a constant multiplying factor. Note that due to the approximation of H_{N_o} used in (44), the latter is a lower bound only for $N_o \geq 2$; nevertheless, making use of the exact expression for H_{N_o}, the right hand side of (44) can be shown to be always lower than (46).

IV. Practical algorithms for secret dither estimation

The theoretical analysis carried out in the previous sections, besides quantifying the information leakage about the secret dither, gives important hints about how to perform dither estimation. Indeed, the information-theoretic formulation given in Section II is closely related to the theory of *set-membership estimation* (SME), aka *set-theoretic estimation* [27], [28], which is widely known in the field of Automatic Control and in certain Signal Processing areas, such as image recovery.[9] In the set-membership formulation of a problem with solution space Ξ, the i-th observation is associated to a subset $\mathcal{F}_i \in \Xi$ that contains all estimates which are consistent with that observation; formally, $\mathcal{F}_i$ can be expressed as

$$\mathcal{F}_i = \{z \in \Xi : \psi_i(z) = 1\},\ i = 1, \ldots, N_o, \tag{47}$$

where $\psi_i(z)$ is a certain indicator function that depends on the problem formulation, and N_o is the number of available observations. The subset $\mathcal{F}$ of estimates which are consistent with all the available information is the so-called *feasible solution set* and is given by $\mathcal{F} = \bigcap_{i=1}^{N_o} \mathcal{F}_i$; finally, a set-membership estimate consists in choosing any point $z \in \mathcal{F}$.

In the dither estimation problem, the solution space of interest is $\mathbb{R}^n$. We will deal for now only with the KMA scenario, deferring until Section IV-C the (minor) modifications needed to cope with the CMA case. Thus, the indicator function is given by

$$\psi_i(\mathbf{z}) = \begin{cases} 1, & \mathbf{z} \in \mathcal{D}_i \\ 0, & \text{otherwise} \end{cases} \tag{48}$$

[9] Interestingly, the set-membership framework has been previously applied to watermark embedding in speech signals [29].

so $\mathcal{F}_i = \mathcal{D}_i$ and $\mathcal{F} = \mathcal{S}_{N_o}$, where $\mathcal{D}_i$ and $\mathcal{S}_{No}$ were defined in Property 1. Moreover, if $\mathbf{T} \sim U(\mathcal{V}(\Lambda))$, which is the worst case for the attacker, the set-membership estimator becomes the maximum likelihood dither estimator. Although intuitively simple, such estimator may not be practical, since exact computation of the solution sets may be computationally prohibitive, because of the increasing number of vertices in $\mathcal{S}_{N_o}$ for $N_o > 1$. Nevertheless, the attacker may not be interested in obtaining the exact $\mathcal{S}_{N_o}$, but instead be satisfied with an accurate approximation of the feasible solution set. Algorithms that are suitable for performing such approximation are discussed in this section. Albeit other algorithms with better performance could be devised, our main purpose is to show that the theoretical information leakage may be exploited in practice with manageable complexity.

According to Property 2, the assumption $\alpha \geq 0.5$ allows us to consider the feasible region as a modulo-Λ convex set. Furthermore, if we shift all observations by $-\tilde{\mathbf{y}}_1 + \mathbf{d}_{m_1}$, then the modulo operation is transparent, so the feasible regions for each observation (Eq. (8)) can be now simplified to[10]

$$\mathcal{D}_i = \tilde{\mathbf{v}}_i + (1-\alpha)\mathcal{V}(\Lambda),\ i = 1, \ldots, N_o, \tag{49}$$

with $\tilde{\mathbf{v}}_i$ defined in (42), rendering the problem convex, since the feasible solution sets (which are in fact polytopes) result from the intersection of convex sets. Some guidelines about how to modify the algorithms in order to work with $\alpha < 0.5$ will be given in Section VIII.

The Voronoi region of any lattice can be described in a variety of ways; for our purposes the most appropriate description is by means of the bounding hyperplanes corresponding to its facets. In the following we assume that, for a Voronoi cell $\mathcal{V}(\Lambda)$ with n_f facets, we know: 1) a vector $\boldsymbol{\phi}_k$ which is outward-pointing normal to the k-th facet; 2) a point $\mathbf{z}_{0,k}$ on the k-th facet. Taking into account each of the modified observations $\tilde{\mathbf{v}}_i$, we have

$$\mathcal{D}_i = \{\mathbf{z} \in \mathbb{R}^n : \boldsymbol{\phi}_k^T(\mathbf{z} - \mathbf{z}_{0,k}) \leq \boldsymbol{\phi}_k^T \tilde{\mathbf{v}}_i,\ k = 1, \ldots, n_f;\ i = 1, \ldots, N_o\}. \tag{50}$$

A. Inner polytope algorithm

The set of modified observations $\{\tilde{\mathbf{v}}_i\}$ together with Eq. (50) define an ensemble of linear inequalities, which in turn describe a polytope in n-dimensional space. Hence, the feasible solution set can be expressed as

$$\mathcal{S}_{N_o} = \left\{\mathbf{z} \in \mathbb{R}^n : \boldsymbol{\phi}_k^T \mathbf{z} \leq \boldsymbol{\phi}_k^T \tilde{\mathbf{v}}_i + \boldsymbol{\phi}_k^T \mathbf{z}_{0,k},\ k = 1, \ldots, n_f;\ i = 1, \ldots, N_o\right\}. \tag{51}$$

We are interested in computing an approximation of the feasible region. For such an approximation to be valid, it must outer bound $\mathcal{S}_{N_o}$ (as tightly as possible), since we do not want to discard any point in $\mathcal{S}_{N_o}$ a priori, and it is also desirable that the approximate region is easy to describe. Then, a reasonable choice is to search for the ellipsoid of minimum volume that contains $\mathcal{S}_{N_o}$ (formally known as the *Löwner-John* ellipsoid of $\mathcal{S}_{N_o}$[30]). Unfortunately, the problem of finding the ellipsoid of interest is ill-posed (indeed, it has been shown to be an NP-complete problem) [31], but on the other hand, the problem of finding the maximum volume ellipsoid contained in the polytope defined by a set of linear inequalities is well-posed. Moreover, if we scale such ellipsoid by a factor of n around its center

[10]Obviously, the offset $-\tilde{\mathbf{y}}_1 - \mathbf{d}_{m_1}$ must be removed from the final estimate.

(n is the dimensionality of the lattice), then the resulting ellipsoid is guaranteed to bound $\mathcal{S}_{N_o}$ [30]. An ellipsoid $\mathcal{E}(\boldsymbol{\theta}, \mathbf{P})$ in Euclidean space is defined by its center $\boldsymbol{\theta}$ and a symmetric positive definite matrix $\mathbf{P}$ such that

$$\mathcal{E}(\boldsymbol{\theta}, \mathbf{P}) = \left\{\mathbf{z} \in \mathbb{R}^n : |(\mathbf{z}-\boldsymbol{\theta})^T \mathbf{P}^{-1} (\mathbf{z}-\boldsymbol{\theta})| \leq 1\right\} = \left\{\mathbf{P}^{1/2}\mathbf{r} + \boldsymbol{\theta} : ||\mathbf{r}|| \leq 1\right\}. \tag{52}$$

The computation of $\hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{P}}$ for the maximum volume ellipsoid contained in $\mathcal{S}_{N_o}$ can be written as a convex minimization problem with second order cone constraints [30]:

$$\begin{aligned} (\hat{\boldsymbol{\theta}}, \hat{\mathbf{P}}) = \arg\min_{\boldsymbol{\theta}, \mathbf{P}} \quad & \log\det(\mathbf{P}^{-1/2}) \\ \text{subject to} \quad & ||\mathbf{P}^{1/2}\boldsymbol{\phi}_k|| \leq \boldsymbol{\phi}_k^T \tilde{\mathbf{v}}_i + \boldsymbol{\phi}_k^T \mathbf{z}_{0,i} - \boldsymbol{\phi}_k^T \boldsymbol{\theta}, \\ & \forall\ k = 1, \ldots, n_f;\ i = 1, \ldots, N_o. \end{aligned} \tag{53}$$

This problem can be recast as a *semidefinite problem* [32] where a linear function is minimized subject to Linear Matrix Inequality (LMI) constraints; this kind of optimization problems can be efficiently solved by means of interior-point methods [31]. As will be checked in Section V, this approach yields tight approximations to $\mathcal{S}_{N_o}$, but it presents an obvious drawback: the potential complexity of the minimization problem arising from the huge number of constraints imposed by large n and N_o. The scheme presented in the next section reduces the complexity by means of an iterative approach.

B. Optimal volume ellipsoid (OVE) [33]

This is a classical SME algorithm that was originally devised for estimation in noisy AR models:

$$y_k = \sum_{j=1}^{n} \theta_j y_{k-j} + u_k = \boldsymbol{\theta}^T \boldsymbol{\phi}_k + u_k,$$

where $\boldsymbol{\phi}_k = (y_{k-1}, \ldots, y_{k-n})^T$ are the n past observations, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$ is the vector of parameters to be estimated, and u_k is the noise term, whose absolute value is assumed to be bounded by γ_k. For the k-th observation, the feasible solution set $\mathcal{F}_k$ is given by all points in $\mathbb{R}^n$ that are *consistent* with the observation, i.e.

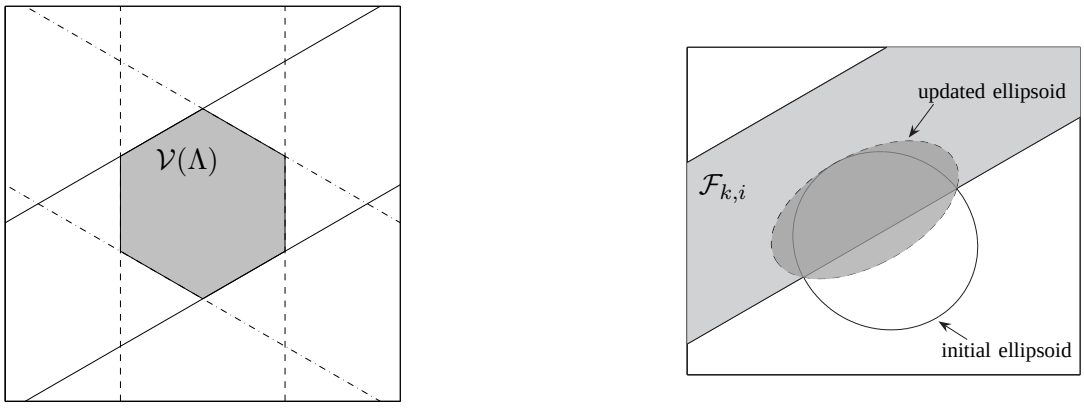
$$\mathcal{F}_k = \{\mathbf{z} \in \mathbb{R}^n : |y_k - \mathbf{z}^T \boldsymbol{\phi}_k| \leq \gamma_k\}. \tag{54}$$

Equation (54) defines a region of $\mathbb{R}^n$ delimited by two parallel hyperplanes:

$$H_{k,1} = \{\mathbf{z} \in \mathbb{R}^n : \mathbf{z}^T \boldsymbol{\phi}_k = y_k - \gamma_k\},\ \ H_{k,2} = \{\mathbf{z} \in \mathbb{R}^n : \mathbf{z}^T \boldsymbol{\phi}_k = y_k + \gamma_k\},$$

which encloses the true parameter vector $\boldsymbol{\theta}$. The series of solution sets is then constructed iteratively as $\mathcal{S}_k = \bigcap_{i=1}^{k} \mathcal{F}_i,\ k = 1, \ldots, N_o$. In order to avoid the costly computation of the exact $\{\mathcal{S}_k\}$, the solution sets are approximately described by means of bounding ellipsoids.

This algorithm can be straightforwardly applied to our problem by slightly modifying the description of the feasible region given in (50): in our case, we need to parameterize $\mathcal{D}_i$ as the intersection of a finite number of parallel hyperplanes. Assuming that the Voronoi cell of the considered lattice is composed of n_f pairwise parallel facets (see



(a) (b)

Fig. 5. (a) Voronoi region of the hexagonal lattice delimited by three pairs of parallel hyperplanes. (b) Intersection between an ellipsoid and a pair of hyperplanes.

Fig. 5-(a)),[11] the feasible solution set for the i-th observation can be specified by a matrix $\mathbf{\Phi}_{n\times n_f/2}$, and a vector $\boldsymbol{\gamma}_{n_f/2\times 1}$ such that $\mathcal{D}_i = \bigcap_{j=1}^{n_f/2} \mathcal{F}_{i,j}$, where

$$\mathcal{F}_{i,j} = \{\mathbf{z} \in \mathbb{R}^n : |\tilde{\mathbf{v}}_i^T \boldsymbol{\phi}_j - \mathbf{z}^T \boldsymbol{\phi}_j| \leq \gamma_j\}, \tag{55}$$

being $\boldsymbol{\phi}_j$ the j-th column of $\mathbf{\Phi}$, and $\gamma_j \triangleq \boldsymbol{\phi}_j^T \mathbf{z}_{0,k}$ is the j-th element of $\boldsymbol{\gamma}$. Hence, the series of solution sets is given by

$$\mathcal{S}_k = \bigcap_{i=1}^{k} \mathcal{D}_i = \bigcap_{i=1}^{k} \bigcap_{j=1}^{n_f/2} \mathcal{F}_{i,j},\ k = 1, \ldots, N_o. \tag{56}$$

The computation of the $(k+1)$-th solution set amounts to obtaining an ellipsoid $\mathcal{E}(\hat{\boldsymbol{\theta}}_{k+1}, \hat{\mathbf{P}}_{k+1}) \supseteq \mathcal{E}(\hat{\boldsymbol{\theta}}_k, \hat{\mathbf{P}}_k) \cap \mathcal{D}_k$. Such ellipsoid is iteratively computed in the following manner:

1) First, make $\mathcal{E}(\mathbf{c}_0, \mathbf{B}_0) = \mathcal{E}(\hat{\boldsymbol{\theta}}_k, \hat{\mathbf{P}}_k)$
2) Compute $\mathcal{E}(\mathbf{c}_{i+1}, \mathbf{B}_{i+1}) \supseteq \mathcal{E}(\mathbf{c}_i, \mathbf{B}_i) \cap \mathcal{F}_{k,i+1},\ i = 0, \ldots, n_f/2 - 1$
3) Finally, make $\mathcal{E}(\hat{\boldsymbol{\theta}}_{k+1}, \hat{\mathbf{P}}_{k+1}) = \mathcal{E}(\mathbf{c}_{n_f/2}, \mathbf{B}_{n_f/2})$

This way, in Step 2 we are intersecting iteratively one ellipsoid with one set $\mathcal{F}_{k,i}$, as is depicted in Figure 5-(b). Clearly, we are interested in finding the ellipsoid with minimum volume that contains such intersection, i.e.

$$\begin{aligned} (\mathbf{c}_{i+1}^*, \mathbf{B}_{i+1}^*) = \arg\min_{\mathbf{c},\mathbf{B}} \quad & \text{vol}(\mathcal{E}(\mathbf{c}, \mathbf{B})) \\ \text{subject to} \quad & \mathcal{E}(\mathbf{c}_i, \mathbf{B}_i) \cap \mathcal{F}_{k,i+1} \subseteq \mathcal{E}(\mathbf{c}, \mathbf{B}). \end{aligned} \tag{57}$$

which is precisely the minimization problem addressed in the OVE algorithm [33], whose analytic solution reads as

$$\mathbf{c}_{i+1}^* = \mathbf{c}_i + \frac{\tau_i \mathbf{B}_i \boldsymbol{\phi}_i}{\left(\boldsymbol{\phi}_i^T \mathbf{B}_i \boldsymbol{\phi}_i\right)^{1/2}}, \qquad \mathbf{B}_{i+1}^* = \delta_i \left(\mathbf{B}_i - \sigma_i \frac{\mathbf{B}_i \boldsymbol{\phi}_i \boldsymbol{\phi}_i^T \mathbf{B}_i}{\boldsymbol{\phi}_i^T \mathbf{B}_i \boldsymbol{\phi}_i}\right), \tag{58}$$

where τ_i, σ_i, δ_i are variables that depend on the observation $\tilde{\mathbf{v}}_k$, the current ellipsoid $\mathcal{E}(\mathbf{c}_i, \mathbf{B}_i)$ and $\mathcal{F}_{k,i+1}$ (details about their calculation can be found in [33]), and finally $\boldsymbol{\phi}_i$ is the i-th column of matrix $\mathbf{\Phi}$.

[11] Should this not be true, the problem can still be recast in a similar manner by adding some additional hyperplanes.

The algorithm just described is obviously optimal in one dimension, since the ellipsoids are simply real intervals. Another interesting feature of this approach, and common to many other iterative SME algorithms, is that further refinements on the solution set are possible by recirculating the observed data, i.e., by feeding to the system the same set of observations repeatedly (as if they were in a circular buffer, for instance). This is possible because the resulting bounding ellipsoid in the i-th iteration depends on both the $(i-1)$-th bounding ellipsoid and the i-th observation. This important feature provides performance similar to that of the above *inner polytope* algorithm, as will be checked in Section V.

C. Dither estimation in the CMA scenario

The CMA scenario implies minor changes to the estimation algorithms proposed above for the KMA case. Actually, estimation in the CMA case can be performed as follows:

1) Assume that the sequence of observations is watermarked with message $\mathbf{m} \in \mathcal{M}$,
2) Perform estimation as in the KMA case,
3) Once $\hat{\mathcal{S}}_{N_o}$ has been obtained, compute the approximate feasible region $\hat{\mathcal{S}}_{N_o}^{CMA}$ as in Eq. (23).
4) Provided that $\mathbf{T} \sim U(\mathcal{V}(\Lambda))$, two possible cases may arise after performing Step 3:

- The resulting feasible regions $(\hat{\mathcal{S}}_{N_o} + \mathbf{d}_m)$ overlap; then, according to Eq. (22), the probability of finding the dither in their intersection is higher than in the remaining regions.
- The regions do not overlap; then, the dither is equally likely in any of the feasible regions.

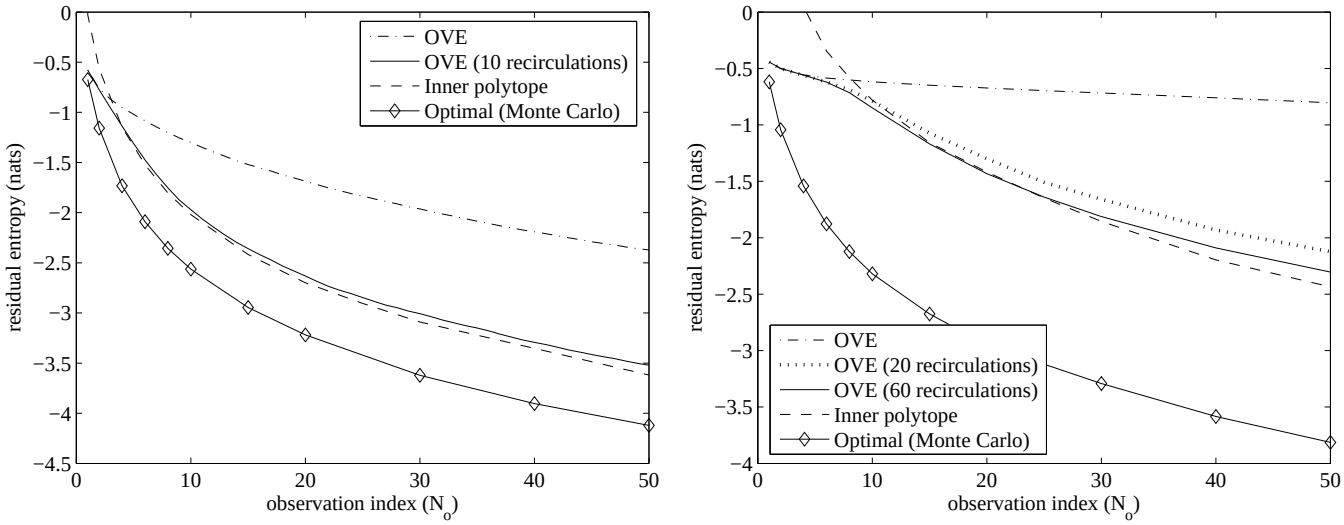
V. Experimental results

This section provides a comparison of the practical performance for the different estimators proposed in Section IV, considering only the KMA scenario. The optimization problems involving LMI's were solved using the optimization packages YALMIP [34] and SeDuMi [35] for Matlab, and the set of observations $\tilde{\mathbf{y}}_i$ was generated according to the distribution given in (5). As for the theoretical part, we will consider here some of the so-called *root lattices* and their duals, introduced in Section III. The Voronoi regions of these lattices are described in [24], from which we derived all the parameters needed for implementing our attack. We provide two different measures of performance of the proposed estimators:

1) the first one is based on the volume of the estimated feasible regions. The volume of the k-th ellipsoid reads as

$$\text{vol}(\mathcal{E}(\hat{\boldsymbol{\theta}}_k, \hat{\mathbf{P}}_k)) = (\det \hat{\mathbf{P}}_k)^{1/2} \cdot V_n(1), \tag{59}$$

where $V_n(1)$ stands for the volume of the n-dimensional sphere of unit radius. When $\mathbf{T} \sim U(\mathcal{V}(\Lambda))$, all points in the interior of the estimated feasible region $\hat{\mathcal{S}}_{N_o}$ have the same probability of being the true dither vector $\mathbf{t}_0$, so it is immediate to compute the residual entropy of the dither as $\log(\text{vol}(\hat{\mathcal{S}}_{N_o}))$. The average value of this *empirical* residual entropy is computed over a large number of realizations. The performance of each method is quantified by the gap between this measure and the theoretical result of Section III.



(a) (b)

Fig. 6. Performance comparison (in terms of residual entropy) for the hexagonal lattice (a) and Gosset lattice E_8 (b), for KMA and $\alpha = 0.5$.

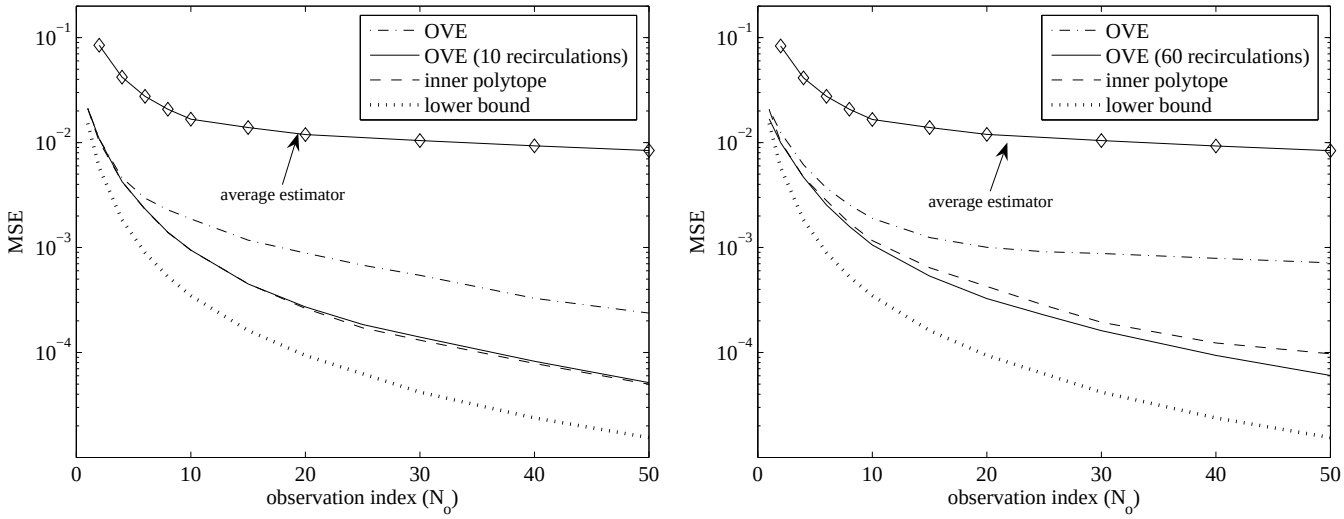
2) the second measure of performance is the squared estimation error per dimension, i.e. $\frac{1}{n}||\mathbf{t} - \hat{\mathbf{t}}||^2$, where $\hat{\mathbf{t}}$ has been taken as the center of the resulting ellipsoid. Note that, as long as this center is close to the center of masses of $\mathcal{S}_{N_o}$, the resulting estimator will be close to the minimum mean-squared error estimator (i.e., the conditional mean estimator). Again, the plots represent this squared error averaged over a large number of observations.

In the experiments, the embedding distortion was fixed to $D_w = \alpha^2/12$, with $\alpha = 0.5$. Figure 6 shows the performance (in residual entropy terms) of the different estimators when the embedding lattices are the hexagonal and E_8 [19]. Although the inner polytope algorithm provides the best performance, it can be observed that the property of recirculation allows to compensate for the loss of optimality of the OVE algorithm. The performance gain is remarkable for the first recirculations, but marginal above a certain number, as can be seen in Fig. 6-(b). Also notice that the number of recirculations must be increased with n in order to match the performance of the inner polytope algorithm. Finally, the plots in Figure 7 show the empirical mean squared error per dimension obtained with each method. The lower bound given by Eq. (40) is plotted for comparison, showing the good performance of both methods. Interestingly, the OVE algorithm seems to perform better than the inner polytope in terms of mean squared error. The performance of the averaging estimator is also plotted for reference; such estimator is optimal for $n \to \infty$ and Λ_n^*, as discussed in Section III-D, but for small n it is clearly far from being so.

A. Possible attacks based on dither estimates

Once the attacker has estimated the dither signal (using the methods proposed here, for instance), he can exploit this knowledge in order to devise powerful attacks against the data hiding scheme which would not be possible for a *blind* attacker. The following are some examples:

1) Complete watermark removal: under the KMA assumptions (i.e., knowledge of the message embedded) the embedding process of lattice DC-DM is fully invertible when the dither is known, as long as the distortion compensation parameter used is smaller than 1 [16, Sect. VII]. This implies that the attacker is able to recover the original host



(a) (b)

Fig. 7. Mean squared error per dimension of the dither estimate, for the hexagonal lattice (a) and Gosset lattice E_8 (b), for KMA and $\alpha = 0.5$.

signal, provided that $\alpha < 1$ and the watermarked signal does not suffer any non-invertible transformation a posteriori, such as clipping or rounding. In the CMA case, there is not a unique possible original host, but the uncertainty is reduced to a finite set of vectors (as many as $|\mathcal{M}|$).

2) Unauthorized embedding of messages: in copy protection scenarios the attacker may remove the watermark inserted in a certain protected content and embed later a different message: for instance, he may change the status of a video from ”Copy Never” to ”Copy Once”.

3) Generation of forgeries: in the authentication scenarios proposed by Eggers *et al.* [17], that are mainly threatened by the CMA attack, as it was discussed in the introduction, the attacker can watermark contents that will be taken as authentic. Notice that for generating a forgery there is no need to know the exact correspondence between messages and coset representatives.

4) And finally, unauthorized decoding of messages embedded in other pieces of content watermarked with the same key. Take into account that reliable decoding is possible only if the dither estimate was obtained in the KMA scenario; in the CMA case, the ambiguity on the embedded message will allow, at most, to check whether different watermarked contents convey the same message or not.

Obviously, the goodness of the host reconstruction in the first attack will depend on the accuracy of the dither estimate at hand. For the other attacks, this accuracy will affect their probability of success, in the sense that poor estimates may lead to the wrong decoding/detection region.

B. Complexity issues

One can find in the literature of set-membership estimation approaches that offer better performance than the ellipsoidal approximations, by computing the exact solution sets [28],[36]. Nevertheless, they may be very computationally demanding in large-scale problems. Instead, the algorithms considered in this paper have proved to be efficient in giving approximate solutions for several hundreds of observations. For the optimization problem in (53), it has been

shown that the number of iterations needed to solve the problem (by means of interior-point methods) does not grow faster than a polynomial of the problem size [32].[12] Most of the computational cost of each iteration lies in the least-squares problem (of the same size as the original problem) that must be solved, whose number of iterations is again polynomial with the problem size. However, in practice it is possible to exploit the problem structure (sparsity, for instance) so as to reduce complexity: in our case, for example, there is a potentially large number of redundant constraints that can be removed for alleviating the computational burden. For high-dimensional lattices it is also possible to simplify the problem description (albeit resulting in looser estimates) by approximating the considered Voronoi region by another simpler polytope that bounds $\mathcal{V}(\Lambda)$.

For the OVE algorithm, the number of arithmetic operations (scalar sums and products) carried out in each iteration is $O(n^2)$. Also, in the OVE algorithm we perform exactly $N_o \cdot \frac{n_f}{2} \cdot n_r$ iterations, where N_o is the number of observations, n_f is the number of facets of the Voronoi cell (equivalently, the number of linear inequalities specifying the problem), and n_r is the number of recirculations of the data. The term n_f will largely depend on the considered lattice, in general, and n_r will be determined by the required accuracy, giving a degree of freedom to the attacker. Finally, it is interesting to note that OVE-like algorithms automatically get rid of redundant constraints, using only those pairs of hyperplanes that produce an update on the solution set.

VI. Comparison: lattice DC-DM vs. Costa

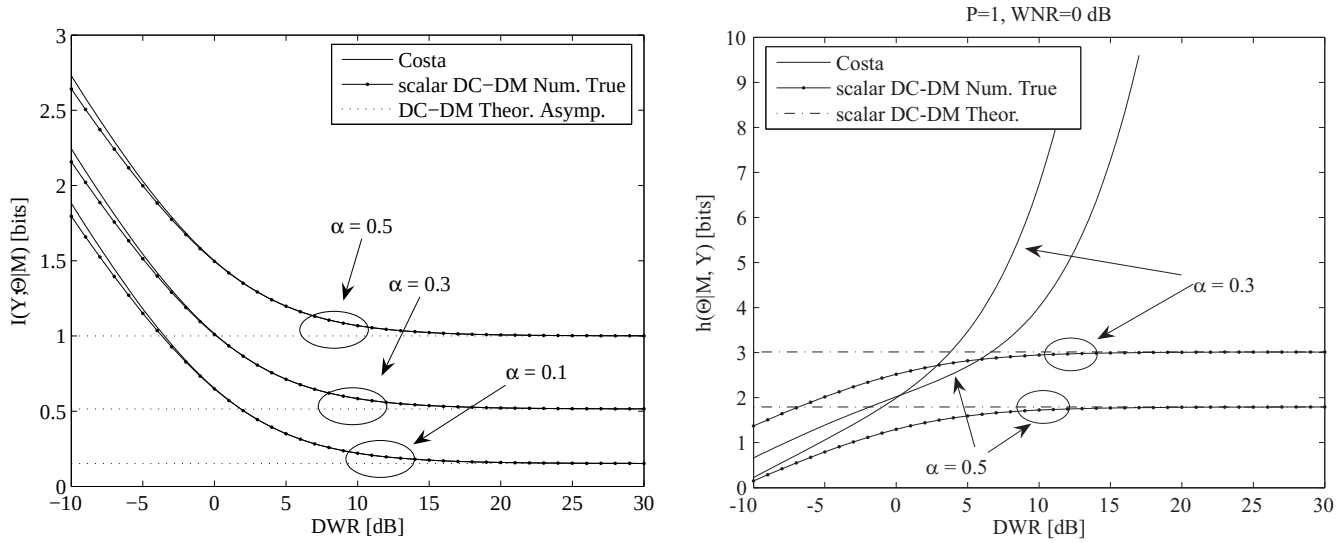
For the lattice DC-DM scheme we have analyzed in Section II, the entropy of the codebook is rather limited due to the codeboook structure and the chosen form of randomization, negatively affecting security. Lattice DC-DM schemes are deeply connected with the theoretical construction developed by Costa [18]. However, the codebook $\mathcal{U}$ in the latter is totally different, since it is random by definition. The main purpose of the brief comparison given in the following is to quantify how much can be gained in terms of security by using a codebook with these characteristics. The theoretical security analysis for Costa's scheme will not be included in this paper due to the lack of space, but it can be found in [10].

In Costa's scheme, for the KMA case and $N_o = 1$, it can be shown that (recall that $h(\mathcal{U}|\mathbf{Y}, M) = h(\mathcal{U}) - I(\mathbf{Y};\mathcal{U}|M)$)

$$\frac{h(\mathcal{U}|\mathbf{Y}, M)}{n} = \frac{h(\mathcal{U})}{n} - \frac{1}{2}\log\left(\frac{P+\sigma_X^2}{(1-\alpha)^2\sigma_X^2}\right), \tag{60}$$

where σ_X^2 and P stand for host and watermark power, respectively, and $h(\mathcal{U})$ denotes the differential entropy of the codebook, given by $h(\mathcal{U}) = \frac{n}{2}|\mathcal{U}| \log\left[2\pi e(P+\alpha^2\sigma_X^2)\right]$. Eq. (60) depends on the ratio $\lambda \triangleq \sigma_X^2/P$ which quantifies the embedding distortion, whereas $|\mathcal{U}|$ depends both on λ and $\xi \triangleq P/\sigma_N^2$, where σ_N^2 is the channel noise. Interestingly, if we make $\lambda \to \infty$ (which corresponds to a low embedding distortion regime), the information leakage for Costa tends to $-n\log(1-\alpha)$, exactly as for DC-DM (see Eq. (17)). Actually, the information leakage in lattice DC-DM also depends on λ, and in fact it is possible to compute this dependency numerically, by means of numerical integration.

[12]The size of an optimization problem is commonly understood as the dimensionality of a vector whose components are the coefficients of the analytical expressions for the constraints and the objective variables.



(a) (b)

Fig. 8. Comparison of the security provided by Costa's scheme and lattice DC-DM, in terms of mutual information (a), and residual entropy (b) per dimension. DWR $\triangleq 10\log_{10}(\lambda)$, and WNR $\triangleq \log_{10}(\xi)$.

In Fig. 8-(a), the information leakage for Costa and scalar DC-DM (i.e., SCS) is shown. It is remarkable the striking similarity in the behavior of both schemes. Furthermore, it can be seen that the asymptotic analysis is in good agreement with the numerical results for the range of embedding distortions of practical interest.

Nevertheless, when the comparison between Costa and lattice DC-DM is made in terms of residual entropy, the similarities disappear (see Figure 8-(b)): whereas for lattice DC-DM the entropy of the codebook is bounded by $\log(\text{vol}(\mathcal{V}(\Lambda)))$, the residual entropy in Costa's scheme is unbounded when $\lambda \to \infty$. The last fact is a consequence of the codebook construction in Costa, where all codewords are mutually independent and its number increases with λ. This constitutes the main advantage, in terms of security, of the random codebook scheme over the lattice scheme that relies solely on dithering. For lattice DC-DM, the number of codewords follow a similar dependence with λ, but every codeword just depends on Λ, the corresponding coset representative, and the secret dither.

On the other hand, for the CMA case, and assuming that the watermarker is transmitting information at the maximum reliable rate allowed by the channel, we have (for $N_o = 1$)

$$\frac{I(\mathbf{Y}_1;\mathcal{U}|\text{CM})}{n} = \frac{I(\mathbf{Y}_1;\mathcal{U}|M)}{n} - \frac{I(\mathbf{Y}_1;M|\mathcal{U})}{n}. \tag{61}$$

This result is clearly related to that given in (24) for DC-DM. Here, we can see that the uncertainty about the codebook increases exactly in the same quantity as the reliable transmission rate.

VII. Application to other scenarios

In this section we discuss the application of the proposed approaches to other related but more involved scenarios. This also shows the importance of the KMA scenario and of the estimators developed for such case.

1) $\alpha < 0.5$: Our analysis was restricted to the case $\alpha \geq 0.5$. In the theoretical part, all the given information leakages constitute upper bounds for $\alpha < 0.5$. For this case, the theoretical analysis gets more intricate, since the feasible region $\mathcal{S}_{N_o}$ may be composed of multiple modulo-Λ convex sets (recall Figure 2). Difficulty of the estimation

is also greatly increased, since it would be necessary to apply several KMA/CMA estimators in parallel, one for each possible convex set. With N_o large enough, all convex feasible regions are likely to vanish except one (in the KMA case), but the increase in the number of convex sets during the first observations may be fairly large, especially when $\alpha \to 0$. In such case, other set-membership approaches suited to non-convex solution sets may perform better [28].

2) Spread Transform Dither Modulation (STDM) [13]: DC-DM schemes may be applied in conjunction with spread transform in low-rate data hiding applications. In that kind of schemes, lattice quantization takes place in a secret projected domain, parameterized by certain projection matrix, and secret dithering can still be used in the projected domain for improving the security of the scheme. Ignorance of the projection matrix invalidates direct application of the estimation algorithms proposed here; however, recent works [2], [12] have shown that independent component analysis (ICA) may be used for estimating the projection matrix. Thus, if ICA is successful, dither estimators may be applied in a second step.

3) Total ignorance of the embedded messages: consider a general scenario where the only information at hand for the attacker is the set of watermarked signals; this is the so-called Watermark Only Attack (WOA), following the nomenclature introduced in [2]. A theoretical analysis similar to that of the KMA may be used to show that in this framework it is possible to achieve (at least theoretically) perfect secrecy in some cases [10], [37], for instance when $\alpha = 0.5$ is used in a binary transmission scheme. In the practical side it is still possible to carry out dither estimation as long as the perfect secrecy condition is not fulfilled; however, KMA estimators cannot be directly applied: one needs to hypothesize first a message sequence and then apply the KMA estimator. However, the problem can be tackled without the need of a brute-force approach if the posterior probability of the message sequences is considered. The maximum likelihood estimate of the message sequence is

$$(\hat{m}_1, \ldots, \hat{m}_{N_o}) = \arg \max_{m_1,\ldots,m_{N_o}} f(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o} | m_1, \ldots, m_{N_o}), \tag{62}$$

and the posterior probability can be factored as

$$f(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o} | m_1, \ldots, m_{N_o}) = \prod_{k=1}^{N_o} \int f(\tilde{\mathbf{y}}_k | m_k, \mathbf{t}) \cdot f(\mathbf{t} | \tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{k-1}, m_1, \ldots, m_{k-1}) d\mathbf{t}, \tag{63}$$

where the conditional pdf of the secret dither is given by (11). When the pdf of the secret dither is uniform, computation of each term in (63) is straightforward, since it is proportional to the volume of $\mathcal{D}_k \cap \mathcal{S}_{k-1}$. Based on this factorization, dither estimation in the WOA scenario may be thought of as a tree search where a KMA estimator is applied to each branch, and each of these branches corresponds to a hypothesized message sequence whose probability can be computed through (63). During the tree search, those branches with low probability may be discarded for simplifying the estimation. Moreover, if the value of α is above a certain threshold (which depends on $|\mathcal{M}|$ and the specific lattice partition) the complexity of the tree search can be dramatically reduced because all branches with non-null probability can be written in terms of a unique branch. As an interesting byproduct of this approach, an estimate of the embedded message sequence can be also obtained. Nevertheless, notice that these are only the main guidelines of the procedure that should be applied to the WOA scenario; a more rigorous and complete analysis will be published elsewhere.

4) Permutations: the security of a lattice DC-DM scheme may be improved by applying secret permutations to the host vectors. This introduces an additional degree of uncertainty that invalidates the direct application of the estimators proposed in this paper. However, if the same permutation is used in multiple watermarked blocks, it is still possible to exploit the information leakage, as shown in the next example: assume that the host is partitioned in l n-length vectors $\mathbf{x}_i$, $i = 1, \ldots, l$, and these vectors are arranged in a $n \times l$ matrix $\mathbf{X}$. Given a secret permutation matrix $\mathbf{P}$, the columns of the new matrix $\mathbf{X}' = \mathbf{PX}$ are watermarked using the standard lattice DC-DM scheme, yielding a watermarked matrix $\mathbf{Y}'$. Later on, the inverse permutation is applied to $\mathbf{Y}'$, obtaining $\mathbf{Y}$, and its rows are the observations that are made available to the attacker. Depending on the symmetry properties of the embedding lattice, two possible cases arise:

1) The lattice is symmetric to permutations of its components. This happens, for instance, to the cubic and *checkerboard* (aka *quincunx*) lattices in 2 dimensions [7], [19]. If this is the case, then the attacker can run the dither estimation algorithm disregarding the actual permutation, obtaining an estimate of the permuted dither. It is easy to see that this permuted estimate allows the same attacks as those discussed in Section V-A, as long as the permutation and the secret dither are the same in the attacked contents.
2) The lattice is not symmetric to permutations. The main consequence is that the feasible regions for the dither are different under each permutation, and this can be exploited to detect inconsistent arrangements in the components of the observations, i.e., those arrangements that produce an empty feasible region cannot be correct. Some experiments performed with the OVE algorithm and the hexagonal lattice have shown that, using 10 recirculations, an average of 32 observations are needed to successfully detect inconsistent arrangements of the components. Using the inner polytope algorithm it is also possible to check inconsistencies: one just needs to run the *feasibility test* to check whether all constraints in the optimization problem can be simultaneously satisfied or not. If not, the considered arrangement is inconsistent.

VIII. Conclusions

The main conclusion of this work is that lattice DC-DM schemes for data hiding relying only on secret dithering are vulnerable to security attacks both in the KMA and CMA scenarios, of practical interest as discussed in the Introduction. For the scenarios considered in this paper, it was shown in Section III that the security level (in terms of residual entropy) can be enlarged by increasing the dimensionality and choosing the appropriate lattice quantizer, although the gain for small n is rather limited; also, asymptotic values are given for the equivocation and the variance of the estimation error, explaining the fundamental gap between the security of DC-DM schemes and spread spectrum methods. Section V shows the strong link between the information-theoretic and set-membership estimation frameworks, applying the latter for the first time to attacks in the data hiding scenario. Additionally, the results in that section confirm that (suboptimal) attacks to security can be made with manageable complexity, yielding accurate dither estimates. This highlights the need for key management solutions, such as those proposed in [38] through temporal redundancy control, in order to reduce the number of observations conveying information about the same dither sample.

The comparison given in Section VI shows that the security weaknesses of lattice DC-DM are not inherent to quantization-based schemes, but they are due to the fact that the randomness of the codeboook relies only on secret dithering. A possible improvement using permutations was briefly considered in Section VII, but dither estimation attacks still seem to be possible, at least with low-dimensional lattices. A new strategy, recently proposed in [39], is the application of secret rotations to the embedding lattice. This approach, in conjunction with permutations, still keeps the structure of the codebook (which is desirable from an implementation point of view) while increasing its a priori entropy. Obviously, the counterpart is the increase needed in the length of the key, but it still constitutes a promising strategy that deserves rigorous analysis in the future.

ACKNOWLEDGMENTS

The authors want to thank Prof. Pierre Moulin for his useful comments that helped to improve the final version of this paper, and also to Tie Liu for helpful discussions about the proof of Theorem 1.

APPENDIX I
RESIDUAL ENTROPIES IN ONE DIMENSION

Here we compute the mean value of (28). It can be seen that $W = 2\mu + \min\{\tilde{V}_1, \ldots, \tilde{V}_{N_o}\} - \max\{\tilde{V}_1, \ldots, \tilde{V}_{N_o}\}$. Hence, $W = 2\mu + X$, where X is the random variable defined as

$$X \triangleq \min\{\tilde{V}_1, \ldots, \tilde{V}_{N_o}\} - \max\{\tilde{V}_1, \ldots, \tilde{V}_{N_o}\},$$

where $x \in (-2\mu, 0]$, so the pdf of W is $f_W(w) = f_X(w - 2\mu)$. This allows us to rewrite the problem as

$$E[\log(W)] = \int_0^{2\mu} \log(w) \cdot f_W(w) dw. \tag{64}$$

First, let us see how the pdf of X can be computed. For having $X = x$, it should be $\min\{\ldots\} = t$ and $\max\{\ldots\} = t - x$; this is so when $\tilde{v}_i = t,\ \tilde{v}_j = t - x$, and $t \leq \tilde{v}_k \leq t - x$, for $k = \{1, \ldots, N_o\} \setminus \{i, j\}$, but taking into account that there are infinite values of t that yield $X = x$. Hence, the pdf of X reads as

$$f_X(x) = N_o(N_o - 1) \int_{-\mu}^{\mu+x} f_{\tilde{V}_i}(t) \cdot f_{\tilde{V}_i}(t - x) \cdot (\text{Prob}\{t < \tilde{V}_i < t - x\})^{N_o - 2} dt, \tag{65}$$

where the factor $N_o(N_o - 1)$ comes from the number of different orderings of the minimum and the maximum in vector $(\tilde{v}_1, \ldots, \tilde{v}_{N_o})$; since all observations are i.i.d., we can simply multiply the integral by this factor. When $\tilde{V}_i \sim U(-\mu, \mu)$, computation of (65) in this case is straightforward and yields

$$f_X(x) = N_o(N_o - 1) \cdot \frac{(-x)^{N_o - 2}}{((1 - \alpha)\Delta)^{N_o}} \cdot [(1 - \alpha)\Delta + x], \tag{66}$$

for $\mu = (1 - \alpha)\Delta/2$. By inserting (66) into (64) and applying integration by parts recursively, the residual entropy results finally in

$$E[\log(W)] = \log(1 - \alpha)\Delta - H_{N_o} + 1, \tag{67}$$

where $H_{N_o} = \sum_{i=1}^{N_o} \frac{1}{i}$ is the N_o-th harmonic number.

APPENDIX II
LOWER BOUND ON THE EQUIVOCATION

By the definition of mutual information, we have

$$I(\tilde{\mathbf{Y}}_2,\ldots,\tilde{\mathbf{Y}}_{N_o};\mathbf{T}'|M_2,\ldots,M_{N_o}) \;=\; h(\tilde{\mathbf{Y}}_2,\ldots,\tilde{\mathbf{Y}}_{N_o}|M_2,\ldots,M_{N_o}) - \sum_{i=2}^{N_o} h(\tilde{\mathbf{Y}}_i|\mathbf{T}',M_i). \tag{68}$$

The first term of (68) can be bounded from above as [6]

$$h(\tilde{\mathbf{Y}}_2,\ldots,\tilde{\mathbf{Y}}_{N_o}|M_2,\ldots,M_{N_o}) = h(\mathbf{Z}_2+\mathbf{T}',\ldots,\mathbf{Z}_{N_o}+\mathbf{T}') \le \sum_{i=1}^{n} h(Z_{i,2}+T_i',\ldots,Z_{i,N_o}+T_i'), \tag{69}$$

where $Z_{i,j}$ is the i-th component of $\mathbf{Z}_j$, and T_i' denotes the i-th component of $\mathbf{T}'$. Since the host signals $\mathbf{X}_j$ and the secret dither $\mathbf{T}$ are mutually independent, it follows that $Z_{i,j}$ and T_i' are independent. Hence, we can write

$$\mathbf{R} \triangleq \text{Cov}(Z_{i,2}+T_i',\ldots,Z_{i,N_o}+T_i') = \mathbf{R}_{Z_i}+\mathbf{R}_{T_i'}, \tag{70}$$

where $\mathbf{R}_{Z_i} \triangleq \text{Cov}(Z_{i,2},\ldots,Z_{i,N_o})$, and $\mathbf{R}_{T_i'} \triangleq \text{Cov}(T_i',\ldots,T_i')$. Furthermore, it follows from Assumption 2 that the self-noise is white [26] with variance per dimension $(1-\alpha)^2P(\Lambda_n^*)$. Hence, by considering that $Z_{i,j}$ are mutually independent for all j, we have

$$\mathbf{R}_{Z_i} = (1-\alpha)^2P(\Lambda_n^*)\cdot\mathbf{I}_{N_o-1}, \qquad \mathbf{R}_{T_i'} = (1-\alpha)^2P(\Lambda_n^*)\cdot\begin{pmatrix} 1 & 1 & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & 1. \end{pmatrix}, \tag{71}$$

for $i=1,\ldots,n$. This allows us to bound Eq. (69) as [6, Th. 9.6.5]:

$$h(\tilde{\mathbf{Y}}_2,\ldots,\tilde{\mathbf{Y}}_{N_o}|M_2,\ldots,M_{N_o}) \;\le\; \frac{n}{2}\log\left((2\pi e)^{N_o-1}|\mathbf{R}|\right) = \frac{n}{2}\log\left((2\pi e(1-\alpha)^2P(\Lambda_n^*))^{N_o-1}\cdot N_o\right). \tag{72}$$

The equivocation or residual entropy is

$$h(\mathbf{T}'|\tilde{\mathbf{Y}}_2,\ldots,\tilde{\mathbf{Y}}_{N_o},M_2,\ldots,M_{N_o}) = h(\mathbf{T}') - I(\tilde{\mathbf{Y}}_2,\ldots,\tilde{\mathbf{Y}}_{N_o};\mathbf{T}'|M_2,\ldots,M_{N_o}), \tag{73}$$

hence, using (68) and (72), Eq. (73) can be lower bounded as

$$h(\mathbf{T}'|\tilde{\mathbf{Y}}_2,\ldots,\tilde{\mathbf{Y}}_{N_o},M_2,\ldots,M_{N_o}) \ge h(\mathbf{T}') + \sum_{i=2}^{N_o} h(\tilde{\mathbf{Y}}_i|\mathbf{T}',M_i) - \frac{n}{2}\log\left((2\pi e(1-\alpha)^2P(\Lambda_n^*))^{N_o-1}\cdot N_o\right). \tag{74}$$

Taking into account that $h(\tilde{\mathbf{Y}}_i|\mathbf{T}',M_i) = h(\mathbf{T}') = h(\mathbf{T}) + n\log(1-\alpha)$, and rearranging terms, we finally arrive at the following lower bound to the equivocation per dimension:

$$\frac{1}{n}h(\mathbf{T}'|\tilde{\mathbf{Y}}_2,\ldots,\tilde{\mathbf{Y}}_{N_o},M_2,\ldots,M_{N_o}) \;\ge\; N_o\frac{h(\mathbf{T})}{n} - \frac{1}{2}\log\left((2\pi eP(\Lambda_n^*))^{N_o-1}\cdot N_o\right) + \log(1-\alpha), \tag{75}$$

and after substituting $\frac{1}{n}h(\mathbf{T}) = \frac{1}{n}\log(\text{vol}(\mathcal{V}(\Lambda_n^*))) = \frac{1}{2}\log\left(\frac{P(\Lambda_n^*)}{G(\Lambda_n^*)}\right)$, we obtain Eq. (34).

APPENDIX III
PROOF OF THEOREM 1

In order to arrive at Eq. (35), we start from the expression

$$\frac{1}{n}h(\mathbf{T}'|\tilde{\mathbf{Y}}_2,\ldots,\tilde{\mathbf{Y}}_{N_o},M_2,\ldots,M_{N_o}) = N_o\frac{h(\mathbf{T})}{n} + N_o\log(1-\alpha) - \frac{1}{n}h(\tilde{\mathbf{Y}}_2,\ldots,\tilde{\mathbf{Y}}_{N_o}|M_2,\ldots,M_{N_o}), \tag{76}$$

which can be straightforwardly obtained by following the reasoning in Appendix II. First, we note that for the sequence of optimum lattice quantizers Λ_n^* we have [26]

$$\lim_{n\to\infty}\frac{h(\mathbf{T})}{n} = \frac{1}{2}\log(2\pi eP(\Lambda_n^*)). \tag{77}$$

On the other hand, we want to prove that the following relation holds:

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}h(\tilde{\mathbf{Y}}_1,\ldots,\tilde{\mathbf{Y}}_{N_o-1}|M_1,\ldots,M_{N_o-1}) &= \lim_{n\to\infty}\frac{1}{n}h(\mathbf{Z}_1+\mathbf{T}',\ldots,\mathbf{Z}_{N_o-1}+\mathbf{T}')\\ &= \frac{1}{2}\log\left((2\pi e(1-\alpha)^2\cdot P(\Lambda_n^*))^{N_o-1}\cdot N_o\right),\end{aligned} \tag{78}$$

with $\mathbf{Z}_i$, $\mathbf{T}'$ independent and uniformly distributed in $(1-\alpha)\mathcal{V}(\Lambda_n^*)$, being $\mathcal{V}(\Lambda_n^*)$ the Voronoi cell of Λ_n^* with second moment per dimension $P(\Lambda_n^*)$. Notice that we have rearranged the observation indices from 1 to N_o-1, for the sake of clarity. We will prove this result by making use of two lemmas.

Lemma 3: Let $\mathbf{Z}$, $\mathbf{T}'$ be two independent random variables uniformly distributed in $(1-\alpha)\mathcal{V}(\Lambda_n^*)$. We have that

$$\lim_{n\to\infty}\frac{h(\mathbf{Z}+\mathbf{T}')}{n} = \frac{1}{2}\log\left(2\pi e(1-\alpha)^2P(\Lambda_n^*)\cdot 2\right). \tag{79}$$

Proof: The entropy power inequality [6] states that

$$e^{\frac{2}{n}h(\mathbf{Z}+\mathbf{T}')} \geq e^{\frac{2}{n}h(\mathbf{Z})} + e^{\frac{2}{n}h(\mathbf{T}')}. \tag{80}$$

Furthermore, we know that [26]

$$\lim_{n\to\infty}\frac{h(\mathbf{Z})}{n} = \lim_{n\to\infty}\frac{h(\mathbf{T}')}{n} = \frac{1}{2}\log\left(2\pi e(1-\alpha)^2P(\Lambda_n^*)\right), \tag{81}$$

so we can write

$$\lim_{n\to\infty} e^{\frac{2}{n}h(\mathbf{Z})} + e^{\frac{2}{n}h(\mathbf{T}')} = 2\cdot e^{\log(2\pi e(1-\alpha)^2P(\Lambda_n^*))} = 2\pi e(1-\alpha)^2P(\Lambda_n^*)\cdot 2 = e^{\frac{2}{n}h(\mathbf{U})}, \tag{82}$$

with $\mathbf{U}\sim\mathcal{N}(\mathbf{0},2(1-\alpha)^2P(\Lambda_n^*)\cdot\mathbf{I}_n)$. Thus, from Eq. (80) we have that

$$\lim_{n\to\infty}\frac{1}{n}h(\mathbf{Z}+\mathbf{T}') \geq \frac{h(\mathbf{U})}{n} = \frac{1}{2}\log(2\pi e(1-\alpha)^2P(\Lambda_n^*)\cdot 2), \tag{83}$$

and we know from Eq. (72) that

$$\frac{h(\mathbf{Z}+\mathbf{T}')}{n} \leq \frac{1}{2}\log\left(2\pi e(1-\alpha)^2P(\Lambda_n^*)\cdot 2\right) \tag{84}$$

for all n. Hence, by combining (83) and (84) the lemma follows. ■

Lemma 4: For $\mathbf{Z}_i$, $\mathbf{T}'$ uniformly distributed in $(1-\alpha)\mathcal{V}(\Lambda_n^*)$, the following result holds

$$\lim_{n\to\infty}\frac{1}{n}h(\mathbf{Z}_m+\mathbf{T}'|\mathbf{Z}_{m-1}+\mathbf{T}',\ldots,\mathbf{Z}_1+\mathbf{T}') = \frac{1}{2}\log\left(2\pi e(1-\alpha)^2P(\Lambda_n^*)\cdot\frac{m+1}{m}\right),\ \text{for } m\geq 1. \tag{85}$$

Proof: We will prove the result by induction. Since it was proven for $m = 1$ in Lemma 3, we will show now that it is true for $m = i$, assuming that it holds for $m \leq i - 1$. Making use of the entropy power inequality and the convexity of $\log(e^x + c)$ in x [40], we can write

$$\frac{2}{n} h(\mathbf{Z}_i + \mathbf{T}' | \mathbf{Z}_{i-1} + \mathbf{T}', \ldots, \mathbf{Z}_1 + \mathbf{T}) \quad \geq \quad \log\left(e^{\frac{2}{n} h(\mathbf{Z}_i)} + e^{\frac{2}{n} h(\mathbf{T}' | \mathbf{Z}_{i-1} + \mathbf{T}', \ldots, \mathbf{Z}_1 + \mathbf{T}')} \right). \tag{86}$$

By using the chain rule for entropies, it can be shown that the following equivocation can be written as

$$h(\mathbf{T}' | \mathbf{Z}_{i-1} + \mathbf{T}', \ldots, \mathbf{Z}_1 + \mathbf{T}') \quad = \quad i \cdot h(\mathbf{Z}_i) - \sum_{j=1}^{i-1} h(\mathbf{Z}_j + \mathbf{T}' | \mathbf{Z}_{j-1} + \mathbf{T}', \ldots, \mathbf{Z}_1 + \mathbf{T}'), \tag{87}$$

and making use of the inductive hypothesis we have that

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} h(\mathbf{T}' | \mathbf{Z}_{i-1} + \mathbf{T}', \ldots, \mathbf{Z}_1 + \mathbf{T}') &= \frac{i}{2} \log\left(2\pi e (1-\alpha)^2 P(\Lambda_n^*)\right) - \frac{1}{2} \sum_{j=1}^{i-1} \log\left(2\pi e (1-\alpha)^2 P(\Lambda_n^*) \cdot \frac{j+1}{j} \right) \\ &= \frac{1}{2} \log\left(2\pi e \cdot \frac{(1-\alpha)^2 P(\Lambda_n^*)}{i} \right). \end{aligned} \tag{88}$$

Thus, if we take limits in (86) we arrive at the following bound:

$$\lim_{n\to\infty} \frac{1}{n} h(\mathbf{Z}_i + \mathbf{T}' | \mathbf{Z}_{i-1} + \mathbf{T}', \ldots, \mathbf{Z}_1 + \mathbf{T}') \geq \frac{1}{2} \log\left(2\pi e (1-\alpha)^2 P(\Lambda_n^*) \cdot \frac{i+1}{i} \right). \tag{89}$$

Note that from the bounding given in (72) and the inductive hypothesis it follows that

$$\lim_{n\to\infty} \frac{1}{n} h(\mathbf{Z}_i + \mathbf{T}' | \mathbf{Z}_{i-1} + \mathbf{T}', \ldots, \mathbf{Z}_1 + \mathbf{T}') \leq \frac{1}{2} \log\left(2\pi e (1-\alpha)^2 P(\Lambda_n^*) \cdot \frac{i+1}{i} \right). \tag{90}$$

Hence, by combining (89) and (90), the lemma follows. ■

Now, using the chain rule for differential entropies we can write

$$\frac{1}{n} h(\mathbf{Z}_1 + \mathbf{T}', \ldots, \mathbf{Z}_{N_o - 1} + \mathbf{T}') = \frac{1}{n} \sum_{i=1}^{N_o - 1} h(\mathbf{Z}_i + \mathbf{T}' | \mathbf{Z}_{i-1} + \mathbf{T}', \ldots, \mathbf{Z}_1 + \mathbf{T}'), \tag{91}$$

and taking the limit when $n \to \infty$, by virtue of Lemma 4, we arrive at the result given in (78). Finally, by combining (76), (77) and (78) we can conclude that

$$\lim_{n\to\infty} \frac{1}{n} h(\mathbf{T}' | \tilde{\mathbf{Y}}_2, \ldots, \tilde{\mathbf{Y}}_{N_o}, M_2, \ldots, M_{N_o}) = \frac{1}{2} \log(2\pi e P(\Lambda_n^*)) - \frac{1}{2} \log(N_o) + \log(1-\alpha),$$

which is the desired result. If we identify now $P(\Lambda_n^*) = D_w / \alpha^2$, then Theorem 1 follows.

REFERENCES

[1] M. Barni, F. Bartolini, and T. Furon, "A general framework for robust watermarking security," *Signal Processing*, vol. 83, no. 10, pp. 2069–2084, October 2003, special issue on Security of Data Hiding Technologies, invited paper.

[2] F. Cayre, C. Fontaine, and T. Furon, "Watermarking security: theory and practice," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, oct 2005.

[3] P. Comesaña, L. Pérez-Freire, and F. Pérez-González, "Fundamentals of data hiding security and their application to spread-spectrum analysis," in *7th Information Hiding Workshop, IH05*, ser. Lecture Notes in Computer Science. Barcelona, Spain: Springer Verlag, June 2005.

[4] A. Kerckhoff, "La cryptographie militaire," *Journal des sciences militaires*, vol. 9, pp. 5–38, January 1883.

[5] C. E. Shannon, "Communication theory of secrecy systems," *Bell system technical journal*, vol. 28, pp. 656–715, October 1949.

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

[7] P. Moulin and R. Koetter, "Data hiding codes," *Proceedings of IEEE*, vol. 93, no. 12, pp. 2083–2126, December 2005.

[8] M. K. Mihçak, R. Venkatesan, and M. Kesal, "Cryptanalysis of discrete-sequence spread spectrum watermarks," in 5^{th} *International Workshop on Digital Watermarking*, F. A. P. Petitcolas, Ed. Noordwijkerhout, The Netherlands: Springer-Verlag, October 2002, pp. 226–246.

[9] G. Doërr and J.-L. Dugelay, "Security pitfalls of frame-by-frame approaches to video watermarking," *IEEE Transactions Sig. Proc., Supplement on Secure Media*, vol. 52, no. 10, pp. 2955–2964, oct 2004.

[10] L. Pérez-Freire, P. Comesaña, and F. Pérez-González, "Information-theoretic analysis of security in side-informed data hiding," in *7th Information Hiding Workshop, IH05*, ser. Lecture Notes in Computer Science. Barcelona, Spain: Springer Verlag, June 2005.

[11] J. Eggers, R. Bäuml, and B. Girod, "Estimation of amplitude modifications before SCS watermark detection," in *Proc. of Security and Watermarking of Multimedia Contents VI*, ser. Proceedings of SPIE, vol. 4675, San Jose, CA, USA, jan 2002.

[12] P. Bas and J. Hurri, "Security of DM quantization watermarking schemes: a practical study for digital images," in *Fourth International Workshop on Digital Watermarking*, M. Barni, I. Cox, T. Kalker, and H. J. Kim, Eds., vol. 3710. Siena, Italy: Springer, September 2005, pp. 186–200.

[13] B. Chen and G. Wornell, "Quantization Index Modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[14] J. Bloom, I. Cox, T. Kalker, J.-P. Linnartz, M. Miller, and C. Traw, "Copy protection for DVD video," *Proc. of the IEEE*, vol. 87, no. 7, pp. 1267–1276, July 1999, Special Issue on Identification and Protection of Multimedia Information.

[15] M. Maes, T. Kalker, J.-P. Linnartz, Joop Talstra, Geert Depovere, and Jaap Haitsma, "Digital watermarking for DVD video copy protection," *Signal Processing Magazine*, vol. 17, no. 5, September 2000.

[16] J. J. Eggers, R. Bäuml, R. Tzschoppe, and B. Girod, "Scalar Costa Scheme for information embedding," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1003–1019, April 2003, special Issue on Signal Processing for Data Hiding in Digital Media and Secure Content Delivery.

[17] J. Eggers and B. Girod, "Blind watermarking applied to image authentication," in *Proc. of Int. Conf. on Audio, Speech and Signal Processing*. Salt-Lake City, USA: IEEE, May 2001.

[18] M. H. M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[19] J. H. Conway and N. J. A. Sloane, *Sphere packings, lattices and groups*, 3rd ed., ser. Comprehensive Studies in Mathematics. New York: Springer-Verlag, 1999, vol. 290.

[20] L. Pérez-Freire and F. Pérez-González, "Spread-spectrum vs. quantization-based data hiding: misconceptions and implications," in *Security, Steganography, and Watermarking of Multimedia Contents VII*, Edward J. Delp III and P. W. Wong, Eds., vol. 5681. San Jose, California, USA: SPIE, January 2005, pp. 341–352.

[21] L. Schuchman, "Dithered signals and their effect on quantization noise," *IEEE Transactions on Communications Technologies*, vol. 12, pp. 162–165, December 1964.

[22] U. Erez and R. Zamir, "Achieving $\frac{1}{2}$log(1+SNR) over the additive white gaussian noise channel with lattice encoding and decoding," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2293–2314, October 2004.

[23] M. Parnas, D. Ron, and R. Rubinfeld, "On testing convexity and submodularity," *SIAM Journal on Computing*, vol. 32, no. 5, pp. 1158–1184, 2003.

[24] J. H. Conway and N. J. A. Sloane, "Voronoi regions of lattices, second moments of polytopes, and quantization," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 211–226, March 1982.

[25] J. Conway and N. Sloane, "Fast quantizing and decoding algorithms for lattice quantizers and codes," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 227–232, March 1982.

[26] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Transactions on Information Theory*, vol. 42, no. 4, pp. 1152–1159, July 1996.

[27] J. R. Deller, "Set membership identification in digital signal processing," *IEEE ASSP Magazine*, vol. 6, no. 4, pp. 4–20, October 1989.

[28] P. L. Combettes, "The foundations of set theoretic estimation," *Proceedings of the IEEE*, vol. 81, no. 2, pp. 182–208, February 1993.

[29] A. R. Gurijala and J. R. Deller, "Speech watermarking with objective fidelity and robustness criteria," in *Conference record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 2, Pacific Grove, CA, USA, November 2003, pp. 1908–1912.

[30] S. Boyd and L. Vandenberghe, *Convex optimization*, ser. SIAM Studies in Applied Mathematics. Cambridge, UK: Cambridge University Press, 2004.

[31] Y. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming*, ser. SIAM Studies in Applied Mathematics. Philadelphia: Society for Industrial and Applied Mathematics, 1994, vol. 13.

[32] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, vol. 38, no. 1, pp. 49–95, March 1996.

[33] M. F. Cheung, S. Yurkovich, and K. M. Passino, "An optimal volume ellipsoid algorithm for parameter set estimation," *IEEE Transactions on Automatic Control*, vol. 38, no. 8, pp. 1292–1296, August 1993.

[34] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004, available from `http://control.ee.ethz.ch/~joloef/yalmip.php`.

[35] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones." *Optimization methods and software*, vol. 11-12, no. 1-4, pp. 625–653, 1999, version 1.1 available from `http://sedumi.mcmaster.ca/`.

[36] E. Walter and H. Piet-Lahanier, "Exact recursive polyhedral description of the feasible parameter set for bounded-error models," *IEEE Transactions on Automatic Control*, vol. 34, no. 8, pp. 911–915, August 1989.

[37] L. Pérez-Freire, F. Pérez-González, and P. Comesaña, "Secret dither estimation in lattice-quantization data hiding: a set-membership approach," in *Security, Steganography, and Watermarking of Multimedia Contents VIII*, Edward J. Delp III and P. W. Wong, Eds. San Jose, California, USA: SPIE, January 2006.

[38] E. Lin and E. Delp, "Temporal synchronization in video watermarking," *IEEE Transactions on Signal Processing*, vol. 52, no. 10, pp. 3007–3022, October 2004.

[39] P. Moulin and A. K. Goteti, "Block QIM watermarking games," *IEEE Transactions on Information Forensics and Security*, September 2006, to appear.

[40] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 279–281, March 1974.