

Joint Iterative Decoding and Estimation for Side-Informed Data Hiding

Félix Balado* *Member, IEEE*, Kevin M. Whelan, Guénolé C.M. Silvestre *Member, IEEE*, and
Neil J. Hurley

Abstract

We present a previously unavailable study on a general procedure for joint iterative decoding and estimation of attack parameters in side-informed data hiding. This type of approach, which exploits iteratively decodable codes for channel identification purposes, has recently become a relevant research trend in many digital communications problems. An advantage is that estimation pilots are not strictly required, thus affording in principle the implementation of blind methods able to work close to the theoretically maximum achievable rate. Such a target naturally requires the use of both near-optimum side-informed data hiding methods (e.g., DC-DM) and near-optimum iteratively decodable channel codes (e.g., turbo codes). The attack channels considered in this study are additive independent random noise, amplitude scaling, and a particular case of fine desynchronization of the sampling grid, whose parameters are estimated by maximum likelihood at the decoder. The complexity of this task is tackled by means of the Expectation-Maximization algorithm, relying on the use of *a priori* probabilities produced by the iterative decoding process.

Index Terms

Side-informed data hiding, additive attacks, amplitude scaling, desynchronization, joint estimation and decoding.

*Corresponding author. E-mail: fiz@ihl.ucd.ie; Phone: +353 1 716 2454; Fax: +353 1 269 7262

All the authors are with the Department of Computer Science at University College Dublin (National University of Ireland). Common address: University College Dublin, Department of Computer Science, Belfield Campus, Dublin 4, Ireland. E-mails: Kevin M. Whelan: kevin.whelan@ihl.ucd.ie; Guénolé C.M. Silvestre: guenole.silvestre@ihl.ucd.ie; Neil J. Hurley: neil.hurley@ucd.ie. Their phone and fax numbers are the same as for the corresponding author. This work was kindly supported by Enterprise Ireland under the research grant ATRP-2002/230 and by the European Commission through the IST Programme under Contract IST-2002-507609 SIMILAR.

I. INTRODUCTION

A crucial watershed in data hiding research has been the realization that, with regard to many important situations, data hiding is just a particular communications problem with side information at the encoder. Nevertheless, the data hiding channel is certainly not a conventional one, as proven by the greater variety of channel distortions (attacks) that may be acceptable—including the intentional ones— compared to customary communications scenarios. In many communications problems the receiver must learn adaptively the channel characteristics. This task, known as channel identification, can be undertaken prior to, or jointly with, the decoding process. In data hiding, channel identification poses an important challenge due to the diversity of transformations that the transmitted (watermarked) signal may undergo. As satisfactory answers to fundamental problems such as the data hiding capacity have already been found, solutions to the identification issue become more necessary than ever. Accordingly, the main objective of this paper is to delve into the identification of the watermarking channel by means of parameter estimation.

Occasionally, identification is unnecessary despite the fact that the channel parameters are not completely specified to the receiver. For instance, in noncoherent detection the communications system is insensitive by construction to certain parameters modifying the channel output—typically a phase delay. In a broad sense, as already suggested in [1], we may analogously deem data hiding schemes invariant to certain transformations to be noncoherent. Examples of this type of data hiding system are some proposals for counteracting coarse desynchronizations [2] or featuring intrinsic amplitude scaling invariance [3], among others. But even if we are able to build useful noncoherent methods such as the preceding ones, it seems clear that there are cases which are not likely to admit this approach. For instance, the design of invariant methods when the involved transformations depend on a relatively large set of parameters is likely to be hampered not only by the complexity of the problem, but also by the fact that noncoherent approaches may decrease the achievable rate, as it actually happens with the aforementioned examples¹.

For these reasons, we may argue that channel identification using parameter estimation will always be necessary to some extent in realistic data hiding implementations. For example, estimation might be used to alleviate the inaccuracies of a possibly inexact noncoherent approach. One possible way to undertake the estimation of channel parameters is with the aid of pilot symbols. These are preestablished training sequences agreed by the encoder and the decoder, and embedded (transmitted) as an overhead

¹During the review of this manuscript we became aware that F. Pérez-González et al. have recently shown that this is not the case for amplitude scaling invariance with their Rational Dither Modulation proposal.

using a regular data hiding (communications) scheme. This shared information facilitates the estimation task of the decoder, sometimes through a careful choice of the symbol sequence. Pilot symbols have been proposed for the estimation of amplitude scaling and fine desynchronization of the sampling grid in side-informed data hiding schemes by Eggers et al. [4] and Baüml et al. [5].

Alternatively, the estimation task may also rely on properties purposely created on the signal released by the encoder. These properties are usually achieved by superimposing a certain pattern on the watermarked signal, sometimes referred to as a template or pilot signal. There are several works that follow this type of approach. Moulin and Ivanović [6] propose to optimize the shape of an additive coarse-synchronization pattern through a game on the fundamental lower bound on the variance of maximum likelihood (ML) estimator. A more extensive analysis in the same spirit is made by the first author in [1] for spread spectrum and quantization-based watermarking, with application to the estimation of cyclic delays and amplitude scalings, respectively. Also, Voloshynovskiy et al. [7] propose watermarks with redundant spatial structure to aid the estimation of coarse and fine random desynchronizations (random bending attack) and the parameters of space-varying additive distortions.

Whereas estimation approaches based on pilot symbols or signals are no doubt useful in practice, they also inherently imply an efficiency loss with respect to the maximum achievable rate. This loss is due to a decrease in the number of conveyable information symbols in the first case, and to a decrease in the available watermark power in the second one, assuming that the perceptual constraints are constant. Moreover, as argued in [1], pilot symbols also mean a loss of adaptivity and security in data hiding, despite the fact that they might be key-dependent. Partly because of these issues, other approaches have investigated blind estimation². Blind methods also carry an efficiency loss, although, ideally, it is only caused by the intrinsic nonzero variance of any estimation process. Among them, a number of so-called content-based blind methods have resorted to heuristic estimation algorithms based on perceptually invariant features of the host signal. Unfortunately, this type of approach presents problems of adaptivity and efficiency, and it makes it difficult to assess performance systematically. A group of works more related to our approach are those by Lee et al. [8] and Shterev et al. [9], who propose methods for blind estimation of amplitude scalings for side-informed data hiding.

Nevertheless, the general blind estimation approach pursued here markedly differs from the strategies followed in the aforementioned works. These former methods were solely based on invariants of the

²Not to be confused with the basic and widespread assumption of blind data hiding, which means that the host signal is unavailable to the receiver.

watermarked signal itself and/or on its statistical properties. Our approach does use this type of information, but it also builds on the redundant structure imposed on the watermark by near-optimal channel coding. Notice that these error-correcting codes create an overhead, similarly to the training sequences. Nonetheless, this overhead is compulsory in any scheme trying to approach channel capacity, and it is asymptotically minimal by definition. Additionally, side-informed methods based on quantizers have been shown to be the most efficient with respect to data hiding capacity. Therefore, we will focus on the study of channel estimation for these watermarking techniques.

The key element of the approach that we will follow here is the iterative refinement of the estimates, afforded by the intertwining of the decoding and estimation processes. The roots of this type of strategy lie in the development of iteratively decodable channel codes (i.e., turbo codes, low-density parity-check codes), with coding gains that enable reliable communication near the Shannon limit [10]. Different researchers in the digital communications field have realized that the so-called *turbo principle* can also be exploited for purposes other than the decoding process strictly speaking. Among these applications we may cite equalization [11], SNR estimation [12], or synchronization [13]. This revolution has motivated a large corpus of research on the subject. The main novelty of the present work lies in the study and application of this philosophy to side-informed data hiding scenarios for the first time.

This paper is organized as follows. In Section II the data hiding framework is presented, and the general iterative estimation and decoding strategy is developed. The methodology in that section is particularized for different scenarios in Section III. Last, the empirical verifications of our proposals are presented in Section IV, followed by the final conclusions.

II. FRAMEWORK AND MODEL

A. Notation and Preliminaries

Scalar random variables are denoted by capital letters, e.g. X , while their realizations are indicated with lowercase types, e.g. x . An exception to this is given by the estimate of a parameter θ , which is denoted by $\hat{\theta}$ for keeping the usual notation in estimation. All vectors are row vectors, and are denoted by bold types. Vectors may be written in capital, e.g. \mathbf{X} , or lowercase letters, e.g. \mathbf{x} , according to the aforementioned convention. The probability distribution function (pdf) of a continuous random variable X is denoted by $p_X(x)$, whereas if X is discrete its probability mass function (pmf) is designated by $P_X(X = x)$. For the sake of simplicity some notational shortcuts will be used. The subscripts of the distribution functions will be dropped wherever it is clear the random variable they refer to. Also, $P_X(X = x)$ will be denoted by $P(x)$ wherever the meaning of this probability is unambiguous. Last, it

will be understood that X represents any of the scalar random variables in $\mathbf{X} = (X_1, \dots, X_N)$ if they are identically distributed.

The host data will be denoted by the N -length random vector \mathbf{X} . Except otherwise indicated, we will limit our exposition to zero-mean Gaussian independent identically distributed (i.i.d.) host data with variance σ_x^2 . This choice allows for analytic tractability in many cases and has been the usual benchmark host in most prior data hiding research. We will also assume without loss of generality that the information symbols to be embedded are statistically independent and equally likely. Two measurements that we will use are the *watermark-to-noise ratio* (WNR) and the *host-to-watermark ratio* (HWR). These parameters are defined as the ratio between the average energy of the involved signals.

B. Near-optimal Scalar Data Hiding

We will briefly recall next the basics of the data hiding approach that we will be considering. Among others, two completely equivalent *scalar* methods have been proposed to approach the capacity limit stated by Costa for communications with side information at the encoder, namely Distortion-Compensated Dither Modulation (DC-DM) with scalar uniform quantizers by Chen and Wornell [14], and the Scalar Costa Scheme (SCS) by Eggers et. al [4]. While relatively simple, the achievable rate of this scalar scheme has been shown to be asymptotically (for large constellation sizes and large WNR) only 1.53 dB away from the data hiding capacity for the Gaussian channel [14], [4]. Notice, that there has been further work to close this gap by means of multidimensional approaches with iterative decoding [15], to which the strategies that we will propose here can be in principle extended along the guidelines provided.

In the remainder we will refer to the aforementioned scalar scheme as DC-DM, due to the precedence of this name. For the sake of simplicity we will consider only a binary scheme, which does not imply any significant loss in the achievable rate with respect to other Q -ary schemes for moderate to low WNR [4]. In any case, all the results that we will present are straightforwardly extendable to higher-dimensional constellations. In binary DC-DM employing scalar uniform quantizers each sample of the watermarked signal carries one information symbol $b_k \in \{\pm 1\}$. This symbol is hidden by quantizing a sample of the host signal x_k to the nearest centroid $Q_{b_k}(x_k)$ of the shifted lattice Λ_{b_k} given by

$$\Lambda_{b_k} \triangleq 2\Delta \mathbb{Z} + \Delta \frac{(b_k + 1)}{2} + d_k,$$

with \mathbb{Z} the only one-dimensional lattice (integer lattice), 2Δ the quantization step, and d a key-dependent pseudorandom dither value deterministically known to both the encoder and the decoder. If we define

next the quantization error with respect to Λ_{b_k} as

$$e_k \triangleq x_k - Q_{b_k}(x_k) = x_k \bmod \Lambda_{b_k}, \quad (1)$$

then, the watermarked signal at the sample k when b_k is embedded is obtained as

$$y_k = x_k - \alpha \cdot e_k = Q_{b_k}(x_k) + (1 - \alpha) \cdot e_k, \quad (2)$$

i.e., the watermark $w_k \triangleq y_k - x_k$ is built using the quantization error weighted by an optimizable constant $0 \leq \alpha \leq 1$. Due to perceptual reasons we have that the inequality $\Delta^2/3 \ll \sigma_x^2$ usually holds true. Then, for a wide range of host signals, $P_{X_k}(x_k)$ can be taken to be roughly constant within one quantization bin. Therefore, we will assume that the quantization error inside a bin is viewed by the decoder as a uniformly distributed random variable in the interval $[-\Delta, \Delta)$. Accordingly, the watermark is also viewed as uniformly distributed in a single bin. Notice that the aforementioned perceptual restrictions usually impose a high HWR = σ_x^2/σ_w^2 .

In uncoded DC-DM the decoder acts by quantizing a received vector \mathbf{z} , which is a distorted version of \mathbf{y} . This sample-by-sample quantization amounts to finding the closest centroid of $\Lambda_{-1} \cup \Lambda_1$ in a Euclidean distance sense. Then, the decoding decisions are

$$\hat{b}_k = \arg \min_{b_k \in \{\pm 1\}} |z_k - Q_{b_k}(z_k)|, \quad (3)$$

for $k = 1, \dots, N$. Last, notice that this particular formulation of DC-DM (SCS) is completely equivalent to those ones which apply a scaling α before quantization.

1) *Channel coding*: In order to make the embedding efficient with respect to the achievable rate, we will hide a binary codeword $\mathbf{c} = (c_1, \dots, c_N)$ in \mathbf{x} instead of hiding N uncoded bits using DC-DM. The codeword is obtained by encoding a binary information vector $\mathbf{b} = (b_1, \dots, b_M)$, $M < N$, using a rate $r = M/N$ error-correcting code. Without loss of generality we will also consider that the codeword symbols are given in antipodal form, i.e., $c_k \in \{\pm 1\}$. We will center our attention on the simplest case of parallel concatenated codes with iterative decoding, i.e., turbo codes, following the exposition in Section I. In any case, the procedures that will be presented here are extensible to other iteratively decodable codes. We recall that, in this case, the parallel concatenated codewords are formed as the concatenation of three elements [10]

$$\mathbf{c} = (\mathbf{c}^s, \mathbf{c}^{p1}, \mathbf{c}^{p2}). \quad (4)$$

The subvector $\mathbf{c}^s = (c_1, \dots, c_M) = \mathbf{b}$ is just the systematic output. On the other hand, the subvectors \mathbf{c}^{p1} and \mathbf{c}^{p2} are the parity output of a certain recursive systematic convolutional (RSC) encoder, and they

have the same length $(N - M)/2$. These parities are obtained when the input to the RSC is \mathbf{b} and a pseudorandomly interleaved version of \mathbf{b} , $\Pi(\mathbf{b})$, respectively. Notice that, without loss of generality and in order to simplify notation, we are assuming that each codeword symbol c_k is embedded in the host sample with the same index x_k , $k = 1, \dots, N$. In practice, a key-dependent pseudorandom permutation may be used for introducing uncertainty in the position of the coded symbols.

C. ML Decoding and Estimation: the Role of the Expectation-Maximization Algorithm

We assume next that, before reaching the receiver, the watermarked signal has undergone a known transformation dependent on an unknown set of deterministic parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_L) \in \Theta \subset \mathbb{R}^L$. We denote this transformation as

$$\mathbf{Z} = T(\mathbf{Y}; \boldsymbol{\theta}), \quad (5)$$

noting that the DC-DM watermarked signal and the received signal are both random variables for the decoder. As seen later, we will consider that this function may also include random transformations depending on the parameters. Next, we will present a general methodology for the estimation of $\boldsymbol{\theta}$ for the framework previously described. This methodology will be particularized for different forms of the transformation (5) in Section III.

Let us assume first that the decoder has received \mathbf{z} and knows the parameter vector $\boldsymbol{\theta}$. Now, thanks to the model $p_{\mathbf{Z}}(\mathbf{z}; \boldsymbol{\theta})$ and to the code structure, the decoder may undertake near-ML decoding instead of just performing symbol-by-symbol hard decisions using (3). This is done by computing the log-likelihood ratio (LLR)

$$\lambda_k \triangleq \log \frac{P(b_k = +1 | \mathbf{z}; \boldsymbol{\theta})}{P(b_k = -1 | \mathbf{z}; \boldsymbol{\theta})}, \quad (6)$$

for $k = 1, \dots, M$. The LLRs are computed iteratively by the decoding process until convergence is achieved. One decoding iteration comprises two semi-iterations consisting in decoding sequentially $(\mathbf{z}^s, \mathbf{z}^{p1})$ and $(\mathbf{z}^s, \mathbf{z}^{p2})$. Notice that we are adopting for \mathbf{z} the same superscript notation used for \mathbf{c} in (4). A decoding semi-iteration computes the *a posteriori* probabilities in (6) using the BCJR algorithm (named after its inventors [16]), which uses the model of \mathbf{Z} and the *a priori* information given by the previous decoding semi-iteration. Therefore, this information, i.e., $\boldsymbol{\lambda}^{(i-1)} = (\lambda_1^{(i-1)}, \dots, \lambda_M^{(i-1)})$, is used for computing $\boldsymbol{\lambda}^{(i)}$, where the superscript (i) indicates the i -th decoding semi-iteration and $\boldsymbol{\lambda}^{(0)} = \mathbf{0}$. After any decoding semi-iteration the decisions on the decoded symbols are simply

$$\hat{b}_k = \text{sgn } \lambda_k, \quad (7)$$

for $k = 1, \dots, M$. After convergence, this procedure amounts to near-ML decoding of a random-like code, accounting for the near-optimal properties of turbo codes. For a detailed description of the iterative decoding process see for instance [10].

Nevertheless, the statistical model needed for this decoding process is dependent on the unknown parameters by hypothesis. Therefore, the receiver has to estimate θ from \mathbf{z} in order to undertake decoding. A commonly favored estimation strategy is the ML approach which, although not optimal for finite data, asymptotically yields the minimum variance unbiased estimator. Following this approach the receiver computes the ML estimate as

$$\hat{\theta}_{\text{ML}} = \arg \max_{\tilde{\theta}} \log p(\mathbf{z}; \tilde{\theta}), \quad (8)$$

where we use the log-likelihood for convenience. For a start, the computation of the likelihood in (8) is hindered by the dependences introduced in \mathbf{Z} by the codeword. For this reason, and being aware of the suboptimality of this strategy, we will resort to solve (8) only for the subvector \mathbf{z}^s of \mathbf{z} corresponding to the systematic part of the codeword. If we assume that, for a given transformation, the samples of \mathbf{Z}^s can be assumed to be independent, then the simpler ML problem can be written as

$$\hat{\theta}_{\text{ML}} = \arg \max_{\tilde{\theta}} \log p(\mathbf{z}^s; \tilde{\theta}) \quad (9)$$

$$= \arg \max_{\tilde{\theta}} \sum_{k=1}^M \log p(z_k; \tilde{\theta}). \quad (10)$$

We recall here that the first M elements of \mathbf{z} are those corresponding to the systematic part of the codeword. In any case, we will see that the subvectors corresponding to the parities \mathbf{z}^{p_1} and \mathbf{z}^{p_2} can be used to improve the estimator (9) beyond what we could obtain with \mathbf{z}^s alone.

Unfortunately, a tractable and straightforward solution to this problem is not possible in many cases. A convenient way to address the problem (10) is provided by the classic Expectation-Maximization (EM) algorithm formalized by Dempster et al. [17], which aims at finding the ML solution iteratively using two alternating steps. The first one, called E-step, involves the computation of the functional

$$\mathcal{Q}(\hat{\theta}^{(i)}, \tilde{\theta}) \triangleq \sum_{\mathbf{b} \in \mathcal{B}} P(\mathbf{b} | \mathbf{z}^s; \hat{\theta}^{(i)}) \cdot \log p(\mathbf{z}^s | \mathbf{b}; \tilde{\theta}), \quad (11)$$

with $\mathcal{B} = \{\pm 1\}^M$, $\hat{\theta}^{(i)}$ the estimate of θ at semi-iteration i , and $\tilde{\theta}$ an optimization variable for obtaining $\hat{\theta}^{(i+1)}$. The embedded information vector \mathbf{b} , unknown to the receiver, plays the role of the *unobserved data* in the terminology employed in [17]. We may see that the computation of (11) requires having a pmf of these unobserved data, conditioned to the received signal \mathbf{z}^s and an estimate of the parameter vector $\hat{\theta}^{(i)}$. This pmf could be computed from \mathbf{z}^s by applying Bayes rule. But actually, the iterative decoding

process does nothing else than obtaining more accurately this distribution by exploiting the codeword redundancy and the *a priori* information. This probability distribution, computed by the turbo decoding process using the estimate of the channel model parameters at a given semi-iteration, is

$$P(b_k|\mathbf{z}; \hat{\boldsymbol{\theta}}^{(i)}) = \frac{1}{1 + \exp(-b_k \lambda_k)}, \quad (12)$$

for $k = 1, \dots, M$, and noting that $\lambda^{(i)}$ depends on \mathbf{z} , $\hat{\boldsymbol{\theta}}^{(i)}$, and $\lambda^{(i-1)}$. For the sake of keeping notation simple, we will denote the pmf (12) just as $P(b_k)$ in the remainder of the paper, keeping always in mind the dependences stated above.

Now, it is possible to take the second step, called M-step, which is the maximization problem

$$\hat{\boldsymbol{\theta}}^{(i+1)} = \arg \max_{\tilde{\boldsymbol{\theta}}} \mathcal{Q}(\hat{\boldsymbol{\theta}}^{(i)}, \tilde{\boldsymbol{\theta}}). \quad (13)$$

It is guaranteed that, for the succession built in this way, $p(\mathbf{z}; \hat{\boldsymbol{\theta}}^{(i+1)}) \geq p(\mathbf{z}; \hat{\boldsymbol{\theta}}^{(i)})$ [17].

An issue is that solving the optimization problem (13) can be involved in some cases. For this reason, it is worth noting that (slower) convergence can be also achieved by finding at each M-step a $\tilde{\boldsymbol{\theta}}$ that increases the functional (11), instead of maximizing it [17]. To sum up, the EM algorithm uses an estimate $\hat{\boldsymbol{\theta}}^{(i)}$ to obtain a better reestimate $\hat{\boldsymbol{\theta}}^{(i+1)}$ that can be used to improve iterative decoding by recomputing (12) and so on. In this way, turbo decoding is intertwined with the estimation problem, as summarized in Figure 1.

As we will discuss more in detail in Section III-C, in some practical situations it is possible to follow a better approach than just updating the model of \mathbf{Z} using $\hat{\boldsymbol{\theta}}^{(i+1)}$ to recompute (12). In those cases, the transformation (5) can be partially reversed using the new estimate $\hat{\boldsymbol{\theta}}^{(i+1)}$ to get $\mathbf{z}^{(i+1)} = T^{-1}(\mathbf{z}^{(i)}; \hat{\boldsymbol{\theta}}^{(i+1)})$, with $\mathbf{z}^{(0)} = \mathbf{z}$, and computing (12) instead as $P(b_k|\mathbf{z}^{(i)}; \hat{\boldsymbol{\theta}}^{(0)})$.

From a practical point of view, the EM functional (11) can be considerably simplified using the hypotheses of mutual independence among the elements of \mathbf{Z}^s and \mathbf{B} , respectively, and noting that $p(z_k|\mathbf{b}; \tilde{\boldsymbol{\theta}}) = p(z_k|b_k; \tilde{\boldsymbol{\theta}})$ for $k = 1, \dots, M$. Then, (11) may be rewritten as

$$\begin{aligned} \mathcal{Q}(\hat{\boldsymbol{\theta}}^{(i)}, \tilde{\boldsymbol{\theta}}) &= \sum_{\mathbf{b} \in \mathcal{B}} \prod_{j=1}^M P(b_j) \cdot \sum_{k=1}^M \log p(z_k|b_k; \tilde{\boldsymbol{\theta}}) \\ &= \sum_{k=1}^M \sum_{b_k \in \{\pm 1\}} P(b_k) \log p(z_k|b_k; \tilde{\boldsymbol{\theta}}) \cdot \sum_{\mathbf{b} \in \mathcal{B}} \prod_{\substack{j=1 \\ j \neq k}}^M P(b_j). \end{aligned}$$

As the summation over all the probabilities of any marginal pmf of $P(\mathbf{b}) = \prod_{k=1}^M P(b_k)$ equals one, we

have finally that

$$\mathcal{Q}(\hat{\boldsymbol{\theta}}^{(i)}, \tilde{\boldsymbol{\theta}}) = \sum_{k=1}^M \sum_{b_k \in \{\pm 1\}} P(b_k) \log p(z_k | b_k; \tilde{\boldsymbol{\theta}}). \quad (14)$$

A number of approaches to iterative ML estimation using the EM algorithm are available in the digital communications literature with slight variants. To the best of our knowledge, the earliest proposal in this sense is that by Kaleh and Vallet [18], prior to the widespread use of iteratively decodable codes. Their analysis is similar to the previous exposition, using a Markov sequence for the unobserved data. Also, Noels et al. apply a similar approach to a synchronization problem using iteratively decodable codes [19] (see the references therein for more analogous works). Last, the EM algorithm has also been proposed previously in a side-informed data hiding context in [8], but without joint estimation and decoding.

As in any iterative method, the initialization of the procedure can be critical for convergence to the global optimum. We will see in Section IV that, whereas the choice of $\hat{\boldsymbol{\theta}}^{(0)}$ is both straightforward and effective for some particular problems, it can be an issue in other cases.

Finally, and as it happens in communications, intertwining estimation and decoding becomes essential in low WNR scenarios for achieving good performance. Accordingly, we will center our attention in this case, for which low channel coding rates will be consequently required. Notice that the estimation of the required parameters can usually be undertaken prior to decoding for high WNR.

D. Blind Performance Assessment

An interesting aside from the preceding exposition is that the actual probability of bit error at the decoder, which is defined as $P_b \triangleq \frac{1}{M} \sum_{k=1}^M \Pr(\hat{b}_k \neq b_k)$, can be blindly estimated in a relatively accurate way at any decoding stage. The decoder can do this by exploiting the information provided by the LLRs (6) together with their statistical properties. There are several approaches available to undertake this estimation, as for instance [20]

$$\hat{P}_b = \frac{1}{M} \sum_{k=1}^M \frac{1}{1 + \exp |\lambda_k|}. \quad (15)$$

Notice that this possibility has been generally neglected in data hiding approaches using turbo codes and attempting to assess performance at the decoder. We will see that this blind performance estimation may be used both to improve the initialization of the iterations and to solve ambiguities that the estimators may present.

III. APPLICATION TO CONCRETE SCENARIOS

In this section we will deal with the specific issues raised by the application of the proposed methodology to different scenarios of interest for DC-DM data hiding.

A. Independent Additive Noise

We consider next the problem of iterative decoding under unknown random additive noise independent of the host. Accordingly, we will assume that the transformation (5) takes in this case the shape

$$\mathbf{Z} = \mathbf{Y} + \mathbf{G}, \quad (16)$$

with \mathbf{G} being i.i.d. noise independent of \mathbf{Y} , with unknown second moment and pdf. Our initial strategy in this problem will be the estimation of the pdf of \mathbf{Z} , as we have seen in the preceding section that the knowledge of this function is indispensable to perform iterative decoding. At a first glance the problem at hand seems different to the one tackled in Section II-C, as it actually involves estimating a function and not a finite set of parameters. Nevertheless, we will see that it is possible to formulate the problem in the required form.

Previous data hiding approaches dealing with the same situation have usually resorted to pilot-based channel estimation —see for instance [7] for the application of ML decoding under independent additive noise with unknown pdf. However, blind communications methods that rely on iterative decoding for estimating the actual pdf are more interesting for our purposes, as for instance the method based on a kernel model used in [21]. Our approach will be similar to the latter, but rooted on the methodology presented in Section II-C and taking advantage of the peculiar characteristics of DC-DM.

First, note that, if the pdf $p_G(g)$ were known, the reliability of a received value z conditioned to the embedded symbol b would be given by

$$p_Z(z|b) = \sum_{q \in \Lambda_b} P(Q_b(X) = q) \cdot p_T(z - q), \quad (17)$$

with $p_T(\cdot)$ defined as the convolution of $p_G(\cdot)$ with a uniform distribution on the interval $[-(1-\alpha)\Delta, (1-\alpha)\Delta)$. The derivation of the expression (17) is straightforward from (2) and (16). While (17) would be the optimal pdf for iterative decoding, the problem posed in this way entails estimating an unknown pdf with support on the whole real range. This problem is inherently involved, even thinking of exploiting the structure induced on the pdf by DC-DM.

A suboptimal but convenient alternative is possible by undertaking decoding using the reliability corresponding to $[z]_b \triangleq z \bmod \Lambda_b$ instead of $p_Z(z|b)$, with the modulo operation defined as in (1).

Whereas the use of (17) would lead to ML decoding, this strategy leads to what is known in the literature as ML *lattice* decoding. The performance loss due to this approach is small if $p_X(x)$ is a slowly varying function. Clearly, if X is uniformly distributed both approaches are approximately equivalent for high HWR. The pdf of $[Z]$ is given by

$$\begin{aligned} p_{[Z]}([z]_b) &= \sum_{q \in \Lambda_b} p_Z([z]_b + q | b) \\ &= \sum_{j \in \mathbb{Z}} p_T([z]_b + 2\Delta j), \end{aligned} \quad (18)$$

where we have used the fact that $\sum_{q \in \Lambda_b} P(Q_b(X) = q) = 1$. Therefore, (18) does not involve the host signal pdf. For notational simplicity, and except otherwise indicated, in the remainder of this section we assume that z is a variable reduced modulo Λ_b .

Notice that the support set of the pdf (18) is restricted to $[-\Delta, \Delta)$. This affords its approximation using a simple but general model based on a finite number of rectangular kernels. We can write this model as

$$p(z; \boldsymbol{\theta}) = \sum_{j=1}^L \theta_j \cdot K(z - (j-1) \cdot \delta + \Delta), \quad (19)$$

with the kernels $K(z)$ defined as

$$K(z) \triangleq \begin{cases} 1/\delta, & 0 \leq z < \delta \\ 0, & \text{otherwise} \end{cases},$$

and $\delta \triangleq 2\Delta/L$. This type of model is usually considered to be nonparametric, although nothing precludes us from seeing it as a parametric one in which the parameter vector $\boldsymbol{\theta} \in \Theta = \{\boldsymbol{\beta} \in [0, 1]^L \mid \sum_{j=1}^L \beta_j = 1\}$ has to be estimated.

Next, we particularize the general estimation methodology for this scenario. To this end, we will find it convenient to rewrite (14) using some definitions that we will establish next. First, we define the interval B_j corresponding to the support set of the j -th kernel in (19), that is,

$$B_j \triangleq [(j-1) \cdot \delta - \Delta, j \cdot \delta - \Delta), \quad (20)$$

with $j = 1, \dots, L$. Using (20) we can define in turn the sets of indices

$$\mathcal{I}_b^j \triangleq \{k \in \{1, \dots, M\} \mid [z_k]_b \in B_j\},$$

with $b \in \{\pm 1\}$, $j = 1, \dots, L$. Now, (14) can be rewritten as

$$\mathcal{Q}(\hat{\boldsymbol{\theta}}^{(i)}, \tilde{\boldsymbol{\theta}}) = \sum_{j=1}^L \sum_{b \in \{\pm 1\}} \sum_{k \in \mathcal{I}_b^j} P(b_k = b) \log \frac{\tilde{\theta}_j}{\delta}. \quad (21)$$

According to (13) we have now to maximize (21), which requires taking into account the restriction on θ that guarantees that (19) is a pdf. To this end, we build the following Lagrangian functional:

$$\mathcal{L}(\hat{\theta}^{(i)}, \tilde{\theta}) \triangleq \mathcal{Q}(\hat{\theta}^{(i)}, \tilde{\theta}) - \rho \left(\sum_{j=1}^L \tilde{\theta}_j - 1 \right).$$

Differentiating with respect to $\tilde{\theta}_l$, and equating to zero to obtain the extreme, we can write

$$\frac{\partial \mathcal{L}(\hat{\theta}^{(i)}, \tilde{\theta})}{\partial \tilde{\theta}_l} = \sum_{b \in \{\pm 1\}} \sum_{k \in \mathcal{I}_b^j} P(b_k = b) \frac{1}{\tilde{\theta}_l} - \rho = 0,$$

for $l = 1, \dots, L$.

In order to solve the Lagrange multiplier ρ we just plug the solution of the equation above into the restriction, obtaining

$$\rho = \sum_{j=1}^L \sum_{b \in \{\pm 1\}} \sum_{k \in \mathcal{I}_b^j} P(b_k = b) = M,$$

where the second equality follows from the fact that $P(\mathbf{b})$ is a pmf, and with M the length of the binary information vector. Therefore, the solution to the M-step (13) is given by the expression

$$\hat{\theta}_j^{(i+1)} = \frac{\sum_{b \in \{\pm 1\}} \sum_{k \in \mathcal{I}_b^j} P(b_k = b)}{M}, \quad (22)$$

for $j = 1, \dots, L$.

In order to gain further insight from (22) we may consider the case in which the decoding decisions are reliable, i.e., the absolute values of (6) are high (cf. (15)). This situation happens when iterative decoding converges to the correct embedded information. In this case, if $P(b_k = -1) \approx 0$ then we have that $P(b_k = 1) \approx 1$, and vice-versa. Interestingly, (19) with the parameters update (22) approximately becomes the normalized histogram of \mathbf{z}^s on the bins B_j , using the hard decisions \hat{b}_k in (7) to make the bin assignment of the corresponding z_k .

The choice of the initial parameter vector $\hat{\theta}^{(0)}$ is straightforward, using the symbol-by-symbol hard decisions (3) that would be made if the received codeword were just considered as uncoded information. These hard decisions are used to make the initial computation of (22). Nevertheless, notice that with this approach only the bins of $p(z; \theta)$ roughly corresponding to $|z| \leq \Delta/2$ can be initialized. In the first iteration all that can be done is setting the remaining bins to a residual uniform nonzero value, and normalizing (19) so that it remains a pdf. The same approach must be taken whenever any estimate $\hat{\theta}_j^{(i)}$ is zero, because these ‘‘impossible values’’ would penalize unacceptably the performance of the iterative decoding. Windowing strategies to cope with this situation are also possible.

1) *Optimization of the Estimator:* We may ask next what is the best number of parameters L for optimizing the performance of the previous approach. It is logical to think that, the closer the estimated pdf is to the true one, the better the behavior of the iterative decoding process will be. We could then think that a possible way to undertake this optimization is by means of the Cramér-Rao bound (CRB). Nevertheless, the variance of the estimator is not enough to determine the optimal L , because a too low number of kernels will lead to a poor approximation of the desired function despite of potentially having a low variance. On the other hand, it is possible to verify that the pdf given by the nonparametric model does not meet the basic regularity condition for the CRB to hold, as it usually happens with models like (19) in which the domain of the pdf for which it is nonzero depends on the parameters (see for instance [22]).

Consequently, we will follow a different approach. One feasible way to assess the accuracy of a nonparametric estimator with respect to the true underlying pdf is given by the integrated mean squared error (IMSE) [23]. This measure of the quality of the estimator is defined as

$$\begin{aligned} \text{IMSE}(L) &\triangleq \int_{-\Delta}^{\Delta} E \left\{ \left(p(z; \hat{\theta}) - p(z) \right)^2 \right\} dz \\ &= \sum_{j=1}^L \int_{B_j} E \left\{ \left(\hat{\theta}_j / \delta - p(z) \right)^2 \right\} dz, \end{aligned} \quad (23)$$

where $p(z)$ is given by (18). As aforementioned, assuming high decoding reliabilities we may assume that the estimator function is approximately a histogram of the systematic part of the received vector on the bins B_j . Then, using the minimum IMSE as a guideline, and under mild assumptions on the true pdf, the asymptotically optimum histogram requires $L \geq (2M)^{1/3}$ [23]. This allows to roughly select the number of parameters without knowledge of the pdf being estimated, but we will see in Section IV-A that further considerations apply.

In order to verify that the IMSE criterion is acceptable for optimizing the decoding performance we will compute next the optimum of (23) for the particular case of additive white Gaussian noise (AWGN). In Section IV-A this optimum will be checked using the empirical behavior of the strategy described in the foregoing section. Firstly, we may write each estimate (22) as $\hat{\theta}_j = S_j/M$, with S_j the random variable representing the bin count at the j -th bin. Following Scott [24], we have that S_j follows a binomial distribution with parameters M and θ_j . This owes to the fact that the bin count is obtained through M independent Bernoulli experiments with probability θ_j . These probabilities are given by the true values of the parameters, computed using (18) as

$$\theta_j = \int_{B_j} p(z) dz, \quad (24)$$

for $j = 1, \dots, L$. After some straightforward operations, and using the moments $E\{S_j\} = M\theta_j$ and $\text{Var}\{S_j\} = M\theta_j(1 - \theta_j)$, (23) can be written as

$$\text{IMSE}(L) = \frac{L}{2\Delta} \left\{ \frac{1}{M} - \left(1 + \frac{1}{M}\right) \sum_{j=1}^L \theta_j^2 \right\} + \int_{-\Delta}^{\Delta} p^2(z) dz. \quad (25)$$

Now, we wish to minimize (25) for the case that \mathbf{G} is an i.i.d. Gaussian noise vector with variance σ_g^2 . In this case, the pdf (18) becomes

$$p(z) = \frac{1}{2(1-\alpha)\Delta} \cdot \sum_{j \in \mathbb{Z}} Q \left(\frac{z - (1-\alpha)\Delta + 2\Delta j}{\sigma_g} \right) - Q \left(\frac{z + (1-\alpha)\Delta + 2\Delta j}{\sigma_g} \right), \quad (26)$$

with $Q(z) \triangleq \frac{1}{\sqrt{2\pi}} \int_z^{\infty} \exp(-z^2/2) dz$. As we may see both in (18) and in the expression above, the aliasing of infinite summands makes the expression of the pdf involved and not easily amenable to analysis.

Nevertheless, notice that (18) is periodic on the lattice $2\Delta\mathbb{Z}$ if we do not enforce the restriction $z \in [-\Delta, \Delta)$. Therefore, following Forney et al. [25], it is possible to obtain the Fourier series representation of $p(z)$ using the dual lattice of $2\Delta\mathbb{Z}$. According to the same authors, if the noise variance is not too low we may approximate (26) by the lower frequency terms of the Fourier series, having in this case

$$p(z) \approx \frac{1}{2\Delta} \left(1 + \eta \cos \left(\frac{\pi z}{\Delta} \right) \right), \quad (27)$$

where

$$\eta \triangleq 2 \text{sinc}(1-\alpha) \exp \left(-\frac{\pi^2 \sigma_g^2}{2\Delta^2} \right).$$

Empirical tests show that this approximation is sufficiently accurate in the low WNR range, for which we have argued at the end of Section II-C that we are posing the estimation problem. Using (27) it is possible to obtain the following closed-form approximation to (24)

$$\theta_j \approx \frac{1}{2\pi} (\xi - 2\eta \sin(\xi/2) \cos((j-1/2)\xi)), \quad (28)$$

with $\xi \triangleq 2\pi/L$. Notice that the minimization of (25) with respect to L requires only the first term. The summation required therein can now be computed using (28) as

$$\sum_{j=1}^L \theta_j^2 = \frac{\xi^2 + \eta^2 \cos \xi}{2\pi\xi}. \quad (29)$$

After replacing (29) in (25), we may differentiate the IMSE with respect to L to find the minimum by equating to zero. Nevertheless, the equation thus obtained is not explicitly solvable. An accurate explicit approximation to the minimum is possible by replacing, before differentiation, the cosine in (29) by the first three terms of its Taylor series expansion around zero, which is afforded by the small value of ξ as

L increases. Then, it can be shown that a closed-form approximation to the optimum number of kernels according to (23) for Gaussian noise is given by

$$L^* \approx \text{nint} \left\{ 2\pi \left(\frac{(M+1)\eta^2}{24\pi} \right)^{\frac{1}{3}} \right\}, \quad (30)$$

where $\text{nint}(\cdot)$ stands for the nearest integer function.

B. Amplitude Scaling

Amplitude scaling is a particularly challenging attack for lattice-based schemes such as DC-DM. The amplitude scaling of a signal watermarked with DC-DM creates a mismatch with respect to the quantization step Δ assumed by the decoder. As a consequence, and without measures to combat this attack, the performance of lattice-based schemes decreases rapidly as the scaling applied departs from unity. There are several previous works focused on estimating amplitude scalings for DC-DM in a blind way. One of these approaches is based on exploiting peaks or periodicities in the pdf of the scaled signals [8], [9]. Although this strategy may be useful for high WNRs, it presents problems for low WNRs where these peaks are not easily discernible. An alternative estimation approach proceeds by conditioning the embedding step size on moments of the host signal which scale with the watermarked signal. This approach was suggested in [4], and put into practice in [26] for local environments.

We will describe next the application of the general methodology in Section II-C to this problem. We will assume that the transformation (5) takes in this case the following shape

$$\mathbf{Z} = \gamma \cdot (\mathbf{Y} + \mathbf{G}),$$

where \mathbf{G} is an i.i.d. Gaussian noise vector with known variance σ_g^2 and independent of \mathbf{X} , and γ a positive unknown scalar that we wish to estimate. A similar scenario has been assumed in other works devoted to the estimation of amplitude scaling [4], [1], although the later work does not assume gaussianity. Joint estimation for the noise variance could also be considered along the guidelines given here. Last, for convenience, we choose a scenario in which scaling is applied after noise addition and not the other way round, as in this case the WNR stays constant after the scaling.

We now particularize (14) for this scenario. The parameter we wish to estimate in this case is just $\theta = \gamma$, and $\Theta = \mathbb{R}^+$. Again, a statistical characterization of Z for a given γ and conditioned to an embedded symbol b is required, i.e., $p(z|b; \gamma)$. Notice first that the ML lattice decoding strategy introduced in Section III-A is not useful here, because, as the shifted lattice Λ_b is scaled by an unknown factor, it is not possible to perform the correct reduction of the variables \mathbf{z} modulo Λ_b .

Then, we will employ here the pdf of Z , which can be straightforwardly obtained from the pdf of $R \triangleq Y + G$. The latter is just given by (17) with $p_G(\cdot)$ a Gaussian pdf. The expression of this pdf is involved by construction, and therefore we will pursue an approximation. For the sake of simplicity, we will assume that the key-dependent dither vector is $\mathbf{d} = \mathbf{0}$, but this dither may be straightforwardly included in the model that we will present without modifying the essentials of our conclusions. In order to obtain the approximation, note that if the pdf of X were uniform and the HWR high, the truncated Fourier series (27) would be a sufficient approximation of $p_R(r|b)$ for $b = -1$ up to a normalization factor. Of course, this approximation is worse for the values close to the limits of the uniform distribution. Similarly, we may think of weighting (27) by the distribution of the host when this is not uniform. Then, weighting that expression by a zero-mean Gaussian pdf with variance σ_x^2 , and normalizing the resulting function to retain a pdf, we may approximate the pdf of R by

$$p_R(r|b) \approx \frac{\exp\left(-\frac{r^2}{2\sigma_x^2}\right)}{\sqrt{2\pi}\sigma_x} \cdot \frac{1 - b \cdot \eta \cos\left(\frac{\pi r}{\Delta}\right)}{1 - b \cdot \eta \exp\left(-\frac{\pi^2\sigma_x^2}{2\Delta^2}\right)}. \quad (31)$$

This model fits well with empirical observations as long as the Fourier approximation is accurate for lattice decoding. An interesting aside is that (31) can be used to quantify analytically the differences between ML decoding and ML lattice decoding, but this problem is out of the scope of this paper. A similar model is proposed in [4] for estimation of amplitude scaling using pilot symbols, although with a definition somewhat more heuristic than the one given here.

Now the scaling of R by a factor γ will yield the pdf sought, that is

$$p_Z(z|b; \gamma) = \frac{p_R(z/\gamma|b)}{\gamma}. \quad (32)$$

Notice that (32) requires knowledge of σ_x^2 . A reasonable way to estimate this variance from the received signal is $\hat{\sigma}_x^2 = \hat{\sigma}_z^2/\gamma^2$, where $\hat{\sigma}_z^2 \triangleq \sum_{k=1}^M z_k^2/M$. We may think of other different estimators of σ_x^2 making use of σ_g^2 ; however, the perceptual constraints strongly limit their potential performance improvements, while, as we will verify next, the proposed estimator allows for some convenient simplifications in the expressions required for the maximization problem. Taking now the logarithm of (32) and disregarding the terms that do not affect the maximization, the EM functional (14) can be written as

$$\mathcal{Q}(\hat{\gamma}^{(i)}, \tilde{\gamma}) = \sum_{k=1}^M \sum_{b_k \in \{\pm 1\}} P(b_k) \cdot b_k \left\{ \exp\left(-\frac{\pi^2 \hat{\sigma}_z^2}{2\tilde{\gamma}^2 \Delta^2}\right) - \cos\left(\frac{\pi z_k}{\tilde{\gamma} \Delta}\right) \right\}, \quad (33)$$

where we have used the aforementioned estimate $\hat{\sigma}_x^2$ and the approximation $\log(1+x) \approx x$. This approximation holds for small values of $|x|$ and, hence, it is accurate in this case as $\eta < 1$ in the range

of validity of the model. Then, the value of $\tilde{\gamma}$ which maximizes (33) is given by the solution to the following expression

$$\frac{\partial \mathcal{Q}(\hat{\gamma}^{(i)}, \tilde{\gamma})}{\partial \tilde{\gamma}} = \sum_{k=1}^M \sum_{b_k \in \{\pm 1\}} P(b_k) \cdot b_k \left\{ \exp\left(-\frac{\pi^2 \hat{\sigma}_z^2}{2\tilde{\gamma}^2 \Delta^2}\right) \cdot \frac{\pi \hat{\sigma}_z^2}{\tilde{\gamma} \Delta} - \sin\left(\frac{\pi z_k}{\tilde{\gamma} \Delta}\right) \cdot z_k \right\} = 0. \quad (34)$$

This equation has to be solved numerically, and it presents multiple roots. Still, it can be verified that its behavior is approximately linear in the region around $\hat{\gamma}^{(i)}$, which would be our initial guess for an iterative numerical solution of (34). So, in order to get an explicit solution, we may approximate the first derivative using the first two terms of its Taylor series about $\hat{\gamma}^{(i)}$. The computation may be further simplified by neglecting the exponential summand in (34), as it tends to zero for high HWR assuming that the scaling factor is close to one. We will see in Section IV-B that the method is only effective within a certain environment around this value. In these conditions, the M-step (13) can be solved as

$$\hat{\gamma}^{(i+1)} \approx \frac{\sum_k \sum_{b_k} P(b_k) \cdot b_k \sin\left(\frac{\pi z_k}{\hat{\gamma}^{(i)} \Delta}\right) \cdot z_k}{\sum_k \sum_{b_k} P(b_k) \cdot b_k \cos\left(\frac{\pi z_k}{\hat{\gamma}^{(i)} \Delta}\right) \cdot \frac{\pi z_k^2}{\hat{\gamma}^{(i)2} \Delta}} + \hat{\gamma}^{(i)},$$

where the summations on k and b_k are over the same values as the corresponding summations in (34). Even in the event that this approximation is not able to exactly reach the M-step maximum, it is usually enough for increasing the EM functional. As remarked in Section II-C, this suffices for the convergence of EM.

1) *Cramér-Rao Bound and Comparison with Pilot Strategies:* We will compute next an approximate analytical expression for the fundamental lower bound on $\text{Var}\{\hat{\gamma}\}$ using the approximation employed in the preceding section. It is straightforward to show that in this case the basic regularity condition for the existence of a valid CRB [22] is satisfied. Obtaining first $p(z; \gamma) = \sum_b p(z|b; \gamma)/2$ we may see that whereas this is an even function of z , $\frac{\partial}{\partial \gamma} \log p(z; \gamma)$ is an odd function, which implies that $\text{E}\left\{\frac{\partial}{\partial \gamma} \log p(z; \gamma)\right\} = 0$. Now, differentiating again $\frac{\partial}{\partial \gamma} \log p(z; \gamma)$ with respect to γ , and using the same approximation on the logarithm as in (33), we have that

$$\frac{\partial^2 \log p(z; \gamma)}{\partial \gamma^2} = \frac{1}{\gamma^2} - \frac{3z^2}{\gamma^4 \sigma_x^2} + \zeta \sin\left(\frac{\pi z}{\gamma \Delta}\right) \frac{2\pi z}{\gamma^3 \Delta} + \zeta \cos\left(\frac{\pi z}{\gamma \Delta}\right) \frac{\pi^2 z^2}{\gamma^4 \Delta^2}, \quad (35)$$

with $\zeta \triangleq \eta^2 \exp(-\pi^2 \sigma_x^2 / (2\Delta^2))$. Next, integrating (35) over $p(z; \gamma)$, and making the simplification $\zeta \approx 0$, which holds true for high HWR values, it is tedious but straightforward to show that the bound may be approximated as

$$\text{CRB}(\gamma) \triangleq -\frac{1}{M} \text{E} \left\{ \frac{\partial^2 \log p(z; \gamma)}{\partial \gamma^2} \right\}^{-1} \approx \frac{\gamma^2}{2M}.$$

It is interesting to compare this approximation to the variance of the estimator of γ using a spread-spectrum pilot signal given in [4], as the authors do not provide an analytical expression for the case of

SCS (DC-DM) pilots. As explained therein, for high HWR and $\text{WNR} > -10$ dB, that variance may be approximated as

$$\text{Var}\{\hat{\gamma}_p\} \approx \frac{\gamma^2 \sigma_x^2}{P \sigma_w^2}, \quad (36)$$

with P the number of pilot symbols used. Note that, although in [4] the amplitude scaling is applied before the additive noise, the analysis is applicable here by considering a noise variance σ_g^2/γ^2 instead of σ_g^2 . In any case, this change in the WNR does not affect (36) under the conditions considered.

As ML estimators are asymptotically efficient, let us assume next that the iterative estimation strategy that we have described is able to approach the CRB. We will see in Section IV-B that this is so for the range of γ where the method converges to the right solution. If we consider now that N is fixed both for the pilot strategy and for our proposal, we may see that, for the same embedding rate $r = M/N = M/(M + P)$ and for the same variance of the estimators, the strategy with pilots yields a lower variance only for

$$r < \frac{1}{1 + 2\sigma_x^2/\sigma_w^2}. \quad (37)$$

As an example, the critical rate marked by (37) is approximately 1/200 for $\text{HWR} = 20$ dB. Take also into account that, even in this case, the information embedded in the pilot-aided case is still uncoded and hence more sensitive to noise addition. This shows that, in the case where convergence is achieved, the joint iterative estimation and decoding strategy with moderate embedding ratios provides a clear advantage over the pilot-aided strategy. Of course, SCS pilots have a much better performance, and we will discuss in Section IV-B how the iterative method and these pilot symbols can be mutually beneficial.

C. Desynchronization

Desynchronization is an important practical issue for watermarking algorithms. Arguably, satisfactory solutions to this problem are not yet available inasmuch as fine desynchronization is concerned. Among previous attempts to tackle fine synchronization recovery by means of estimation we may cite [5] and [6], where pilot symbols and signals, respectively, are proposed to aid in the estimation of fine sampling grid desynchronizations.

We will focus our study on a particular case of fine desynchronization. In this scenario, a unidimensional host signal whose samples are spatially contiguous is watermarked using DC-DM. The watermarked signal is then distorted by AWGN, ideally interpolated and resampled at a constant offset with respect to the original sampling grid. A slight variant of the same setting was previously considered for studying the capacity decrease of DC-DM under desynchronization [5]. Although apparently simple, and as shown in that work, this case is already considerably detrimental for DC-DM performance under minor

desynchronization. Assuming that the vector samples come from critically sampling a corresponding continuous signal with sampling period T_s , the transformation (5) takes the shape

$$\mathbf{Z} = (\mathbf{Y} + \mathbf{G}) \cdot \mathbf{H}(\tau). \quad (38)$$

The $N \times N$ matrix $\mathbf{H}(\tau)$ represents an ideal interpolation filter plus resampling at a signed fraction τ of the sampling period. Hence, its elements are defined as $h_{j,k}(\tau) \triangleq \text{sinc}(T_s(k - j + \tau))$. Border effects in the interpolation operation are neglected, as their importance is small for N large. In this case the unknown parameter to estimate is $\theta = \tau$ with $\Theta = \mathbb{R}$, as our working hypothesis will be that the variance of \mathbf{G} is known. This assumption has also been made in [6] for studying pilot signals in the related desynchronization problem of translation, but, again, joint estimation of this variance could also be considered.

Using the same definition of R as in the previous section, one single sample of the received vector can be put now as

$$Z_k = R_k \cdot \text{sinc}(\tau) + \sum_{j \neq k} R_j \cdot h_{j,k}(\tau) \quad (39)$$

$$= R_k \cdot \text{sinc}(\tau) + V_k, \quad (40)$$

where V_k stands for the intersymbol interference (ISI) caused on R_k by the resampling process. As the elements of \mathbf{R} are independent, the ISI term \mathbf{V} may be modeled as i.i.d. AWGN independent of \mathbf{R} . Assuming that the transformation (38) preserves the energy of \mathbf{R} , we have that $\sigma_v^2 = \sigma_r^2 \cdot (1 - \text{sinc}^2(\tau))$ where σ_r^2 is the variance of \mathbf{R} . Notice that (40) resembles the amplitude scaling scenario described in the foregoing section, but we will see that the previous approach is not applicable here.

As in the preceding sections, the estimation procedure described in Section II-C is applied next to this context. With regard to the statistical characterization of \mathbf{Z} required, notice that now the hypothesis of independence between the elements of \mathbf{Z}^s used to obtain (14) is no longer true. This is due to the dependences evident in (39), introduced during the interpolation process. However, for small values of τ these dependences are weak and we will assume that (14) approximately holds, which constrains the applicability of our proposal to this situation. Notice that, in this case, we also have that $\text{sinc}(\tau) \approx 1$. As it will be seen in Section IV-B, an amplitude scaling in the vicinity of unity has little impact on coded DC-DM, what allows us to assume that the fine desynchronization (38) simply amounts to $\mathbf{Z} \approx \mathbf{R} + \mathbf{V}$. Although this scenario is admittedly rather restrictive, it is a fact that even remarkably small resampling offsets can cause probabilities of decoding error close to $1/2$ —in spite of using powerful channel codes—, as we will confirm in Section IV-C. This drastic performance decrease is due to the relatively large value

of σ_v^2 , and, consequently, it is interesting to see what the proposed methodology can achieve under these conditions.

Recalling Section III-A and the previous discussion on the scaling factor, in this setting is possible to resort to lattice decoding. Again, for notational simplicity and except otherwise indicated, we assume that z is a variable reduced modulo Λ_b for the remainder of this section. Then, for AWGN, the pdf of Z may be approximated using (27), taking into account the fact that the variance of the noise is now $\sigma_g^2 + \sigma_r^2 \cdot (1 - \text{sinc}^2(\tau))$ instead of just σ_g^2 . For the range of τ considered, this expression of the variance can be simplified by using the first order Taylor approximation $\text{sinc}^2(\tau) \approx 1 - \pi^2\tau^2/3$ about $\tau = 0$. Then, we may write the required pdf as

$$p(z; \tau) \approx \frac{1}{2\Delta} \left(1 + \eta \cdot \exp \left(-\frac{\pi^4 \sigma_r^2 \tau^2}{6\Delta^2} \right) \cos \left(\frac{\pi z}{\Delta} \right) \right). \quad (41)$$

In practice, for high HWR values it is sufficient to estimate the variance required in (41) from the received signal as $\hat{\sigma}_r^2 = \hat{\sigma}_z^2$, where $\hat{\sigma}_z^2$ is defined as in the preceding section. Observe from the pdf (41) that the estimation problem poses in this case a problem of identifiability, as τ can be only estimated up to a sign factor. Taking the logarithm of (41), the EM functional (14) is given by

$$\mathcal{Q}(\hat{\tau}^{(i)}, \tilde{\tau}) = \sum_{k=1}^M \sum_{b_k \in \{\pm 1\}} P(b_k) \cdot \log \left(1 + \eta \cdot \exp \left(-\frac{\pi^4 \hat{\sigma}_z^2 \tilde{\tau}^2}{6\Delta^2} \right) \cos \left(\frac{\pi [z_k] b_k}{\Delta} \right) \right), \quad (42)$$

where we have disregarded the factors that do not affect the maximization of (42) with respect to $\tilde{\tau}$. The value that maximizes this expression is given by the solution to the following equation, which must be solved numerically

$$\frac{\partial \mathcal{Q}(\hat{\tau}^{(i)}, \tilde{\tau})}{\partial \tilde{\tau}} = \sum_{k=1}^M \sum_{b_k \in \{\pm 1\}} P(b_k) \cdot \left\{ \sec \left(\frac{\pi [z_k] b_k}{\Delta} \right) + \eta \cdot \exp \left(-\frac{\pi^4 \hat{\sigma}_z^2 \tilde{\tau}^2}{6\Delta^2} \right) \right\}^{-1} = 0, \quad (43)$$

where $\sec(x) = 1/\cos(x)$. It can be verified that (43) has just two symmetric roots, and hence $\hat{\tau}^{(i+1)}$ has two possible values except when zero is a double root. According to the block diagram in Figure 1 the model should be updated next using the square of $\hat{\tau}^{(i+1)}$ and a new iteration performed. As only the square of the parameter is needed, no sign ambiguity exists for updating the model (41) and performing a new iteration. In fact, we could have arrived at an identical solution by posing the problem for $\tilde{\xi} \triangleq \tilde{\tau}^2$ instead of $\tilde{\tau}$.

Nevertheless, and as already discussed in Section II-C, whereas this strategy was adequate for the problems studied in Sections III-A and III-B, it is suboptimal in this case because it does not take measures to correct the lack of synchronization. Actually, it is just a noise variance estimation that only allows for a minor refinement of the channel model. A more powerful practical approach entails

resynchronizing \mathbf{z} by using all the available knowledge at each iteration, but this operation requires determination of the sign of $\hat{\tau}^{(i+1)}$. This ambiguity can be resolved by means of the blind performance assessment method described in Section II-D, which allows to estimate the sign of $\hat{\tau}^{(i+1)}$ yielding the lowest estimated probability of decoding error. If the sign is determined, the received vector can be resynchronized at each iteration as

$$\mathbf{z}^{(i+1)} = \mathbf{z} \cdot H\left(-\sum_{j=0}^{i+1} \hat{\tau}^{(j)}\right).$$

This resynchronization is made from $\mathbf{z}^{(0)} = \mathbf{z}$ instead of $\mathbf{z}^{(i)}$ —without any substantial complexity increase—in order not to accumulate the border effect of successive reinterpolations. Last, the reliabilities for the next iteration are computed as described in Section II-C.

It can be verified that the lack of local identifiability about $\tau = 0$ originates a singularity in $\text{CRB}(\tau)$ on that value. In principle, an alternative constrained Cramér-Rao bound could be pursued using the aforementioned energy preservation condition, but we do not deal with this issue here due to space restrictions.

IV. EXPERIMENTAL RESULTS

In this section we undertake empirical tests for the different scenarios analyzed in Section III, using the probability of decoding error as the performance measurement. The component codes of the turbo codes used in the experiments have been chosen by trial-and-error, without extensive optimizations. Unless otherwise indicated, a pseudorandom interleaver of size $M = 1000$ is chosen. A maximum of 20 decoding iterations are performed, and the cross-entropy criterion [27] is used for early stopping of the process. The value of the DC-DM parameter α is chosen empirically to correspond to the approximate WNR at which the turbo cliff for a particular code occurs on the Gaussian channel.

A. Independent Additive Noise

We show in Figure 2 the performance of the method in Section III-A in front of Gaussian noise. The value of $\hat{\boldsymbol{\theta}}^{(0)}$ is set following the initialization procedure explained therein. We can see that, for the values of L close to the minima of the IMSE shown in Figure 3, the iterative estimation and decoding method only loses about 0.2 dB with respect to the case where the channel model is known by the decoder. These minima are correctly given by (30), which confirms the usefulness of the IMSE criterion. For the value of M considered, the selection of the number of parameters independently of the pdf would be given by $L = 13$ —according to the lower bound in Section III-A.1—, which is a good choice.

Also, comparing Figures 2 and 3, we see that for $L > L^*$ the performance is eventually worse than for $L < L^*$, whereas the IMSE is higher in the second case for the range of WNR considered (cf. $L = 16, 32$ with $L = 4$, for $\text{WNR} > 2$ dB). This seemingly paradoxical behavior is explained by the fact that, as L increases, the likelihood of empty bins in the histogram-like nonparametric model also increases. As discussed in Section III-A, zero values have to be modified in order to avoid detrimental zero probabilities for the decoder, causing the loss of performance observed. As we can see in Figure 2, this effect is logically more evident as the WNR increases. Consequently, conservative values for the pdf-independent choice of L are advisable.

B. Amplitude Scaling

We may see in Figure 4 the performance of the scheme in Section III-B. For comparison purposes, the performances of the uncoded case and of the turbo-coded case with no estimation are also included in the plot. The initialization of the method is done using $\hat{\gamma}^{(0)} = 1$. In this case the algorithm achieves the same performance as for the case where the scaling is known by the decoder ($\gamma = 1$) for $\gamma \in (0.95, 1.05)$. An additional performance improvement is also offered over the remaining range of γ considered in the plot. The sudden performance degradation observed outside of this range is due to the fact that the initial LLRs (6) become increasingly less reliable due to the inaccurate channel model obtained when setting $\hat{\gamma}^{(0)} = 1$, thus precluding the iterative method to converge. While the improvement obtained is modest compared with that of the turbo code acting alone, we see in the same figure that further improvements are possible through a judicious initialization of the EM algorithm.

This alternative initialization strategy is based on exploiting the blind performance assessment (15). Starting from $\hat{\gamma}^{(0)} = 1$, a number of decoding iterations are performed and \hat{P}_b is then examined. If the decoding decisions are reliable the algorithm is let to run until convergence; alternatively, a reinitialization is done. Two values of γ either side of the current estimate $\hat{\gamma}^{(i)}$ are chosen, and one decoding semi-iteration is performed using both new values. Again, we employ \hat{P}_b to determine the reinitialization value, which is used to run the algorithm until convergence. This simple strategy, which comes at a minimal increase in complexity, enlarges the range of scalings which the algorithm can correct, as shown in Figure 4.

Although the improvement seems small, notice that the obvious applicability of the method would be the refinement of a coarse and inexpensive pilot-assisted estimation of γ . For example, using the pilot-based approach in [4] it is possible to roughly determine the number of pilot symbols for that estimator to have a variance within the region where the iterative method converges to $\hat{\gamma}_{\text{ML}}$. Using just as an

indication a uniform distribution on $(-0.9, 1.1)$, the results in [4] suggest that less than 500 SCS pilots may be enough to achieve the equivalent target $\delta_{\Delta} \approx 0.05$ in the notation used in that work, and, hence, to get convergence with our iterative method using $\hat{\gamma}^{(0)} = \hat{\gamma}_p$.

Last, Figure 5 shows the empirical variance of $\hat{\gamma}$ compared with the Cramér-Rao lower bound. Notice that the bound is approximately achieved in the range around unity corresponding to near-errorless decoding in Figure 4, and that the analytical bound is reasonably good.

C. Desynchronization

Figures 6 and 7 show the performance results for the method described in Section III-C, for two different HWR values. The algorithm is initialized using $\hat{\tau}^{(0)} = 0$. First, a number of decoding iterations are performed just updating the channel model as in Figure 1, and without attempting resynchronization. This affords a more accurate sign resolution on the critical initial estimate $\hat{\tau}^{(1)}$. Afterwards, the iterative method is run as explained in Section III-C, performing resynchronization based on the updated estimates. It can be seen in Figures 6 and 7 that, in the case in which iterative estimation without resynchronization is undertaken, the performance is similar to that of the turbo code acting alone. This is not surprising as, although the estimate of the overall noise variance is correct, the increasingly accurate estimate of τ is not exploited to remove the ISI noise by resynchronizing. On the other hand, the method with resynchronization achieves near-errorless decoding for offsets $\tau \in (-0.08, 0.08)$ (Figure 6). This is roughly the range for which the assumptions in Section III-C hold, and as we see, it extends substantially the one achieved by use of channel model updates only. As in the case studied in the preceding section, a sudden performance degradation occurs when the initial reliability values (6) are not accurate enough to allow convergence. Also, observe that the improvement afforded by the method is smaller in Figure 7, as the independence hypothesis become less reasonable when the HWR increases. Similarly to the discussion in the preceding section, the use of a coarse estimate obtained with a limited number of pilot symbols could also be used in this case for initialization purposes. Last, Figure 8 shows the empirical variance of the estimator as a function of τ .

V. CONCLUSIONS

We have presented a study on a general estimation approach for side-informed data hiding using iterative decoding, showing its practical application to various relevant scenarios. To the best of our knowledge, this is the first work within the data hiding field that takes this type of approach. Despite of the fact of having used rather controlled environments, our experiments show the potentiality of these kinds of techniques

for data hiding. The main limitations of an iterative approach like the one presented here are clearly given by the initialization issues. Another issue, common to both parameter estimation and invariant building methods and not studied here, is the decision on which attacks to consider when designing the data hiding system. On the positive side, we must realize that the strategy proposed is inherently afforded in many situations, as it does not require anything else than the near-optimal iteratively decodable codes present in most advanced systems. Therefore, it is almost always integrable with other solutions like pilot estimation to achieve overall better system performance. Moreover, the generality of the procedure allows its potential extension to many other situations involving estimation.

Only constant amplitude scaling and desynchronization have been contemplated. Nevertheless, and considering that the Cramér-Rao bound decreases as $O(M^{-1})$, we might think of applying the estimation methods presented to amplitude scalings or desynchronizations only locally constant, for sufficiently large local environments. In addition, notice that the resynchronization method presented in Section III-C is straightforwardly extensible to dimensions higher than one. Last, other emerging approaches for iterative decoding and synchronization in digital communications [13] —as for example, the use of digital PLLs in the iterative process— are also promising for side-informed data hiding scenarios.

ACKNOWLEDGMENT

The authors would like to thank Dr. Fernando Pérez González for suggesting the use of the Expectation Maximization algorithm and Dr. Sviatoslav Voloshynovskiy for interesting discussions. Many thanks also to the associate editor Dr. Pierre Moulin and the anonymous reviewers for valuable suggestions and comments.

REFERENCES

- [1] P. Moulin, “Embedded-signal design for channel parameter estimation,” in *Procs. of the IEEE Workshop on Statistical Signal Processing*, St Louis, USA, September 2003, parts I & II.
- [2] C.-Y. Lin, M. Wu, J. A. Bloom, I. J. Cox, M. L. Miller, and Y. M. Lui, “Rotation, scale, and translation resilient watermarking for images,” *IEEE Trans. Image Processing*, vol. 10, no. 5, pp. 767–782, May 2001.
- [3] H. Malvar and D. Florêncio, “Improved spread spectrum: a new modulation technique for robust watermarking,” *IEEE Trans. Signal Processing*, vol. 51, no. 4, pp. 898–905, 2003.
- [4] J. J. Eggers, R. Bäuml, R. Tzschoppe, and B. Girod, “Scalar Costa scheme for information embedding,” *IEEE Trans. Signal Processing*, vol. 51, no. 4, pp. 1003–1019, April 2003.
- [5] R. Bäuml, J. J. Eggers, and J. Huber, “A channel model for watermarks subject to desynchronization attacks,” in *Procs. of SPIE: Security and Watermarking of Multimedia Contents IV*, vol. 4675, San José, USA, January 2002.

- [6] P. Moulin and A. Ivanović, “The Fisher information game for optimal design of synchronization patterns in blind watermarking,” in *Procs. of IEEE International Conf. on Image Processing*, vol. 2, Thessaloniki, Greece, September 2001, pp. 550–553.
- [7] S. Voloshynovskiy, F. Deguillaume, S. Pereira, and T. Pun, “Optimal adaptive diversity watermarking with channel state estimation,” in *Proc. of SPIE*, ser. Security and Watermarking of Multimedia Contents III, vol. 4314, San José, USA, January 2001, pp. 673–685.
- [8] K. Lee, D. S. Kim, T. Kim, and K. A. Moon, “EM estimation of scale factor for quantization-based audio watermarking,” in *Procs. of the 2nd International Workshop on Digital Watermarking*, ser. Lecture Notes in Computer Science, vol. 2939, Seoul, Korea: Springer-Verlag, 2004, pp. 316–327.
- [9] I. Shterev, R. Lagendijk, and R. Heusdens, “Statistical amplitude scale estimation for quantization-based watermarking,” in *Procs. of SPIE, Security, Steganography, and Watermarking of Multimedia Contents VI*, vol. 5306, San José, USA, January 2004.
- [10] C. Berrou, A. Glavieux, and P. Thitimajshima, “Near Shannon limit error-correcting coding and decoding: Turbo codes,” in *Proc. IEEE Int. Conf. on Communications*, Geneva, Switzerland, May 1993, pp. 1064–1070.
- [11] C. Douillard, M. Jézéquel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, “Iterative correction of intersymbol interference: turbo equalization,” *European Trans. Telecommun.*, vol. 6, pp. 507–511, September–October 1995.
- [12] T. A. Summers and S. G. Wilson, “SNR mismatch and online estimation in turbo decoding,” *IEEE Trans. Communications*, vol. 46, no. 4, pp. 421–423, April 1998.
- [13] J. R. Barry, A. Kavčić, S. W. McLaughlin, A. Nayak, and W. Zeng, “Iterative timing recovery,” *IEEE Signal Processing Mag.*, January 2004.
- [14] B. Chen and G. W. Wornell, “Quantization index modulation: A class of provably good methods for digital watermarking and information embedding,” *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.
- [15] U. Erez and S. ten Brink, “Approaching the dirty paper limit for canceling known interference,” in *41th Ann. Allerton Conf. on Communications, Control and Computing*, Illinois, USA, October 2003.
- [16] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, “Optimal decoding of linear codes for minimizing symbol error rate,” *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 284–287, March 1974.
- [17] A. Dempster, N. Laird, and D. Rubin, “Maximum-likelihood from incomplete data via the EM algorithm,” *J. Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.
- [18] G. K. Kaleh and R. Vallet, “Joint parameter estimation and symbol detection for linear or nonlinear unknown channels,” *IEEE Trans. Communications*, vol. 42, no. 7, pp. 2406–2413, July 1994.
- [19] N. Noels, C. Herzet, A. Dejonghe, V. Lottici, and L. Vandendorpe, “Turbo synchronization: an EM algorithm interpretation,” in *IEEE Intl. Conf. on Communications*, Anchorage, USA, May 2003.
- [20] P. Hoeher, I. Land, and U. Sorger, “Log-likelihood values and Monte Carlo simulation - Some fundamental results,” in *Proc. 2nd Int. Symp. on Turbo Codes & Rel. Topics*, Brest, France, September 2000, pp. 43–46.
- [21] Y. Li and K. H. Li, “Iterative PDF estimation and decoding for CDMA systems with non-Gaussian characterization,” *IEE Electronics Letters*, vol. 36, no. 8, pp. 730–731, April 2000.
- [22] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993, vol. I.
- [23] G. R. Terrell and D. W. Scott, “Oversmoothed nonparametric density estimates,” *Journal of the American Statistical Association*, vol. 80, no. 389, pp. 209–214, March 1985.
- [24] D. W. Scott, “On optimal and data-based histograms,” *Biometrika*, vol. 66, no. 3, pp. 605–610, 1979.

- [25] G. D. Forney, M. D. Trott, and S.-Y. Chung, "Sphere-bound-achieving coset codes and multilevel coset codes," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 820–850, May 2000.
- [26] J. Oostveen, T. Kalker, and M. Staring, "Adaptive quantization watermarking," in *Procs. of SPIE, Security, Steganography, and Watermarking of Multimedia Contents VI*, vol. 5306, San José, USA, January 2004.
- [27] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 429–445, March 1996.

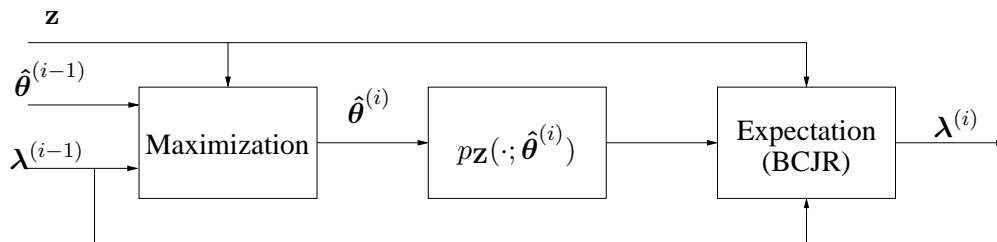


Fig. 1. One semi-step of iterative turbo decoding intertwined with one step of the iterative EM algorithm. For simplicity, necessary interleavings/deinterleavings of \mathbf{z}^s and λ before BCJR are not explicit in the figure.

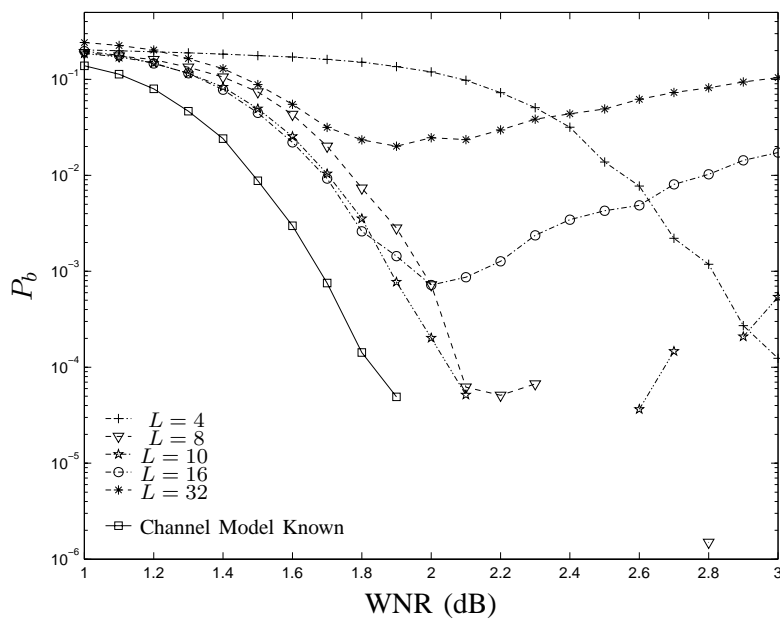


Fig. 2. Performance of DC-DM under additive white Gaussian noise using iterative channel estimation. $M = 1000$, $r = 1/3$, $\alpha = 0.65$, HWR = 25 dB.

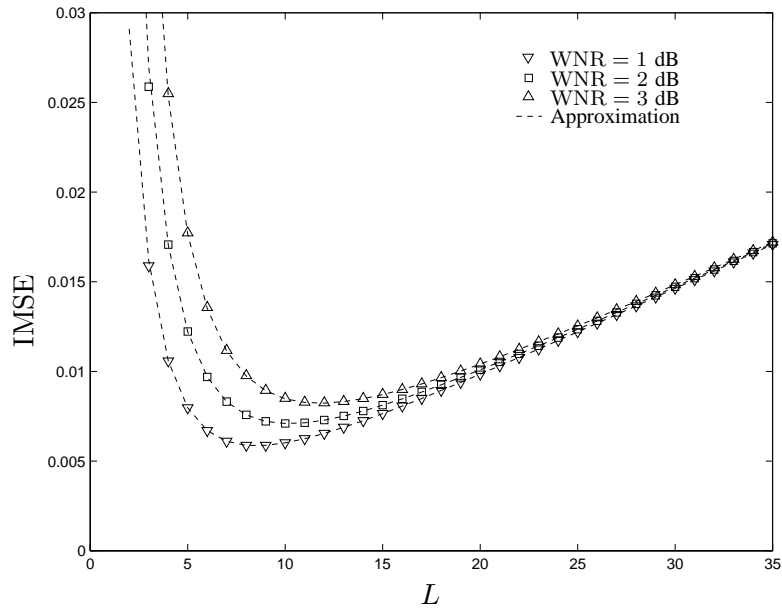


Fig. 3. IMSE vs number of kernels L in the nonparametric model. Gaussian noise, $M = 1000$, $\alpha = 0.65$.

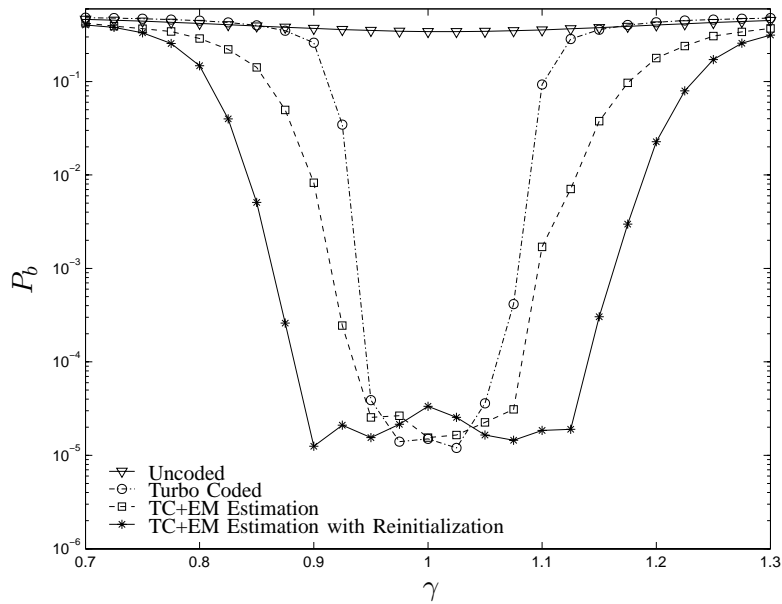


Fig. 4. Performance of DC-DM under an amplitude scaling γ . $r = 1/15$, HWR = 20 dB, WNR = -4 dB, $\alpha = 0.3$, $\hat{\gamma}^{(0)} = 1$.

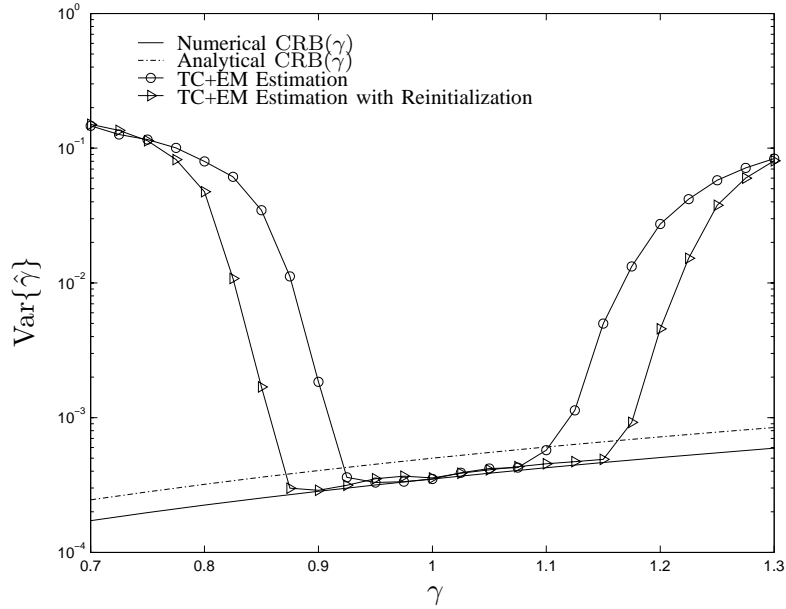


Fig. 5. Variance of the estimator of the amplitude scaling γ . $r = 1/15$, HWR = 20 dB, WNR = -4 dB, $\alpha = 0.3$, $\hat{\gamma}^{(0)} = 1$.

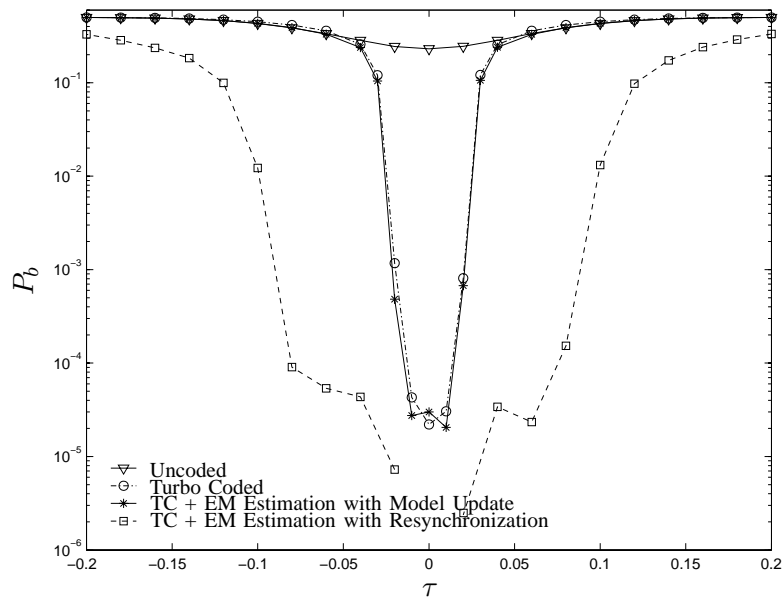


Fig. 6. Performance of DC-DM under fine desynchronization with a constant fraction τ of the sampling period, HWR = 20 dB. $r = 1/5$, WNR = 0 dB, $\alpha = 0.5$, $\tau^{(0)} = 0$.

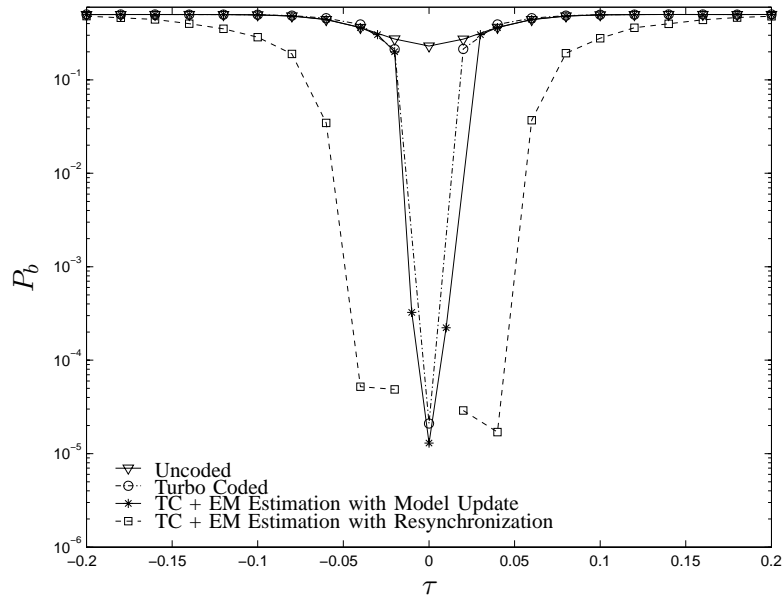


Fig. 7. Performance of DC-DM under fine desynchronization with a constant fraction τ of the sampling period, HWR = 25 dB. $r = 1/5$, WNR = 0 dB, $\alpha = 0.5$, $\tau^{(0)} = 0$.

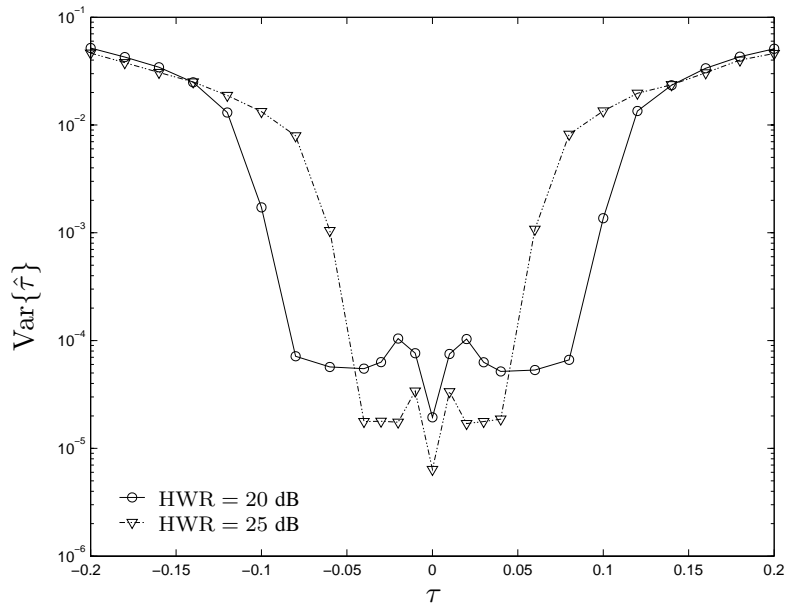


Fig. 8. Variance of the estimator of the desynchronization factor τ using resynchronization. $r = 1/5$, WNR = 0 dB, $\alpha = 0.5$, $\tau^{(0)} = 0$.