Broadcast- and MAC-Aware Coding Strategies for Multiple User Information Embedding

Abdellatif Zaidi, Pablo Piantanida, and Pierre Duhamel, *Fellow, IEEE*

***Abstract*—Multiple user information embedding is concerned with embedding several messages into the same host signal. This paper presents several implementable dirty-paper-coding (DPC)-based schemes for multiple user information embedding, through emphasizing their tight relationship with conventional multiple user information theory. We first show that depending on the targeted application and on whether the different messages are asked to have different robustness and transparency requirements or not, multiple user information embedding parallels one of the well-known multiuser channels with state information available at the transmitter. The focus is on the Gaussian broadcast channel (GBC) and the Gaussian multiple access channel (GMAC). For each of these channels, two practically feasible transmission schemes are compared. The first approach consists in a straightforward—rather intuitive—superimposition of DPC schemes. The second consists in a joint design of these DPC schemes. This joint approach heavily relies on a recent work by Kim *et al.* in which the authors extend the single-user Costa's DPC to the multiple user case. The results in this paper extend the practical implementations quantization index modulation (QIM), distortion-compensated QIM (DC-QIM), and scalar Costa scheme (SCS) that have been originally conceived for one user to the multiple user case. After presenting the key features of the joint design within the context of structured scalar codebooks, we broaden our view to discuss the framework of more general lattice-based (vector) codebooks and show that the gap to full performance can be bridged up using finite dimensional lattice codebooks. Performance evaluations, including bit error rates (BERs) and achievable rate region curves are provided for both methods, illustrating the improvements brought by a joint design.**

***Index Terms*—Broadcast channel (BC), communication with side information, dirty paper coding (DPC), information embedding, lattices, multiple access channel (MAC).**

I. Introduction

RESEARCH on information embedding has gained considerable attention during the last years, mainly due to its potential application in multimedia security. Digital watermarking and data hiding techniques, which are a major branch of information embedding, refer to the situation of embedding information carrying signals called *watermarks* into another signal, generally stronger, called *cover* or *host* signal. The cover signal is any multimedia signal. It can be either image, audio, or video. The embedding must not introduce perceptible distortions to the host, and the watermark should survive common channel degradations. These two requirements are often called *transparency requirement* and *robustness requirement*, respectively. Being conflicting, these two requirements, together with the interference stemming from the host signal itself, have for long time limited the use of digital watermarking to applications where little information (payload) has to be embedded. These include copyright protection [2], for example, where the transmission of just one bit of information, expected to be detectable with very low probability of false alarm, is sufficient to serve as an evidence of copyright. In these applications, the watermark is, in general, a pseudonoise sequence obtained by means of conventional spread-spectrum modulations (SSM) techniques. SSM techniques do not allow the encoder to exploit knowledge of the host signal in the design of the transmitted codewords and are consequently interference limited by construction.

Information embedding can also be viewed as power-limited communication over a "super"-channel with state (or side) information noncausally known to the transmitter [3], [4]. The channel input is the watermark and the available state information is the cover or host signal itself. An achievable rate, for a watermarking system, consists in any rate of payload that can be successfully decodable. The capacity, or more precisely the data hiding capacity, is the supremum of all achievable rates. Based on this equivalence, many host-interference rejecting schemes have been proposed [3], [5] in this still emerging field. It has then become possible to embed large amount of information while at the same time satisfying the two previous requirements. The most relevant work in this area is the initial Costa's "Writing on Dirty Paper" [6], commonly known as "Costa's problem." Costa was the first to examine the Gaussian dirty paper problem. He obtained the remarkable result that an additive Gaussian interference which is noncausally known only at the encoder incurs no loss of capacity, relative to the Gaussian interference-free channel. The theoretical proof of "Costa's problem" is based on an optimal random binning argument for independent identically distributed (i.i.d.) Gaussian codebook. This technique had been proved to be optimal for more general problems in "coding for channels with random parameters" studied in [7] and [8]. Binning consists in a probabilistic construction of codewords. However, this probabilistic construction is convenient only for theoretical analysis, not for practical coding applications. The schemes proposed by Chen and Wornell [3] and Eggers *et al.* [5], in the context of information embedding, adhere to Costa's

Manuscript received March 27, 2006; revised September 5, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Gerald Schuller. The material in this paper was published in part at the 2005 and the 2006 IEEE International Conferences on Acoustics, Speech and Signal Processing. This work was supported by the RNRT under the Projects Secured Diffusion of Music on mObiles (SDMO) and European Network of Excellence for Cryptology (ECRYPT).

A. Zaidi is with the Communications and Remote Sensing Laboratory (TELE), Université catholique de Louvain (UCL), B-1348 Louvain-la-Neuve, Belgium (e-mail: zaidi@tele.ucl.ac.be).

P. Piantanida and P. Duhamel are with the Signals and Systems Laboratory (LSS/CNRS), 91192 Gif-sur-Yvette Cedex, France (e-mail: Pablo.Piantanida@lss.supelec.fr; Pierre.Duhamel@lss.supelec.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2007.893973

setting in that the interference due to the host signal is nearly removed, thus achieving close to the side-information capacity. In addition, these schemes are feasible in practice, because random codewords are replaced by low-complexity quantization-based algebraic codewords. These two sample-wise schemes are referred to as "quantization index modulation" (QIM) and "scalar Costa scheme" (SCS), respectively.

During the last years, both QIM and SCS have been thoroughly studied and extended into different directions such as non-Gaussian channel noise [9], nonuniform quantizers [10], partial state information [11], [12], and recently, lattice codebooks [13]–[17]. This paper extends these schemes to another direction: multiple information embedding.[1] Multiple information embedding refers to the situation of embedding several messages into the same host signal, with or without different robustness and transparency requirements. Of course, finding a single unifying mathematical analysis to general multiple information embedding situations under broad assumptions seems to be a hard task. Instead, this paper addresses the very common situations of multiple user information embedding, from an information theoretic point-of-view. The basic problem is that of finding the set of rates at which the different watermarks can be simultaneously embedded. This problem has tight relationship, as well as in the case of single embedding, to conventional multiple user information theory. Consider, for example, watermark applications such as copy control, transaction tracking, broadcast monitoring, and tamper detection. Obviously, each application has its own robustness requirement and its own targeted data hiding rate. Thus, embedding different watermarks intended to different usages into the same host signal naturally has strong links with transmitting different messages to different users in a conventional multiuser transmission environment. The design and the optimization of algorithms for multiple information embedding applications should then benefit from recent advances and new findings in multiuser information theory [19]. For instance, in this paper, we first argue that many multiple information embedding situations can be nicely modeled as communication over either a broadcast channel (BC) or a multiple access channel (MAC), both with state information available at the transmitter(s). Next, we rely heavily on the general theoretical solutions for these channels (cf. [19]) to devise efficient practical encoding schemes. The resulting schemes consist, in essence, of applying the initial QIM or SCS as many times as the number of different watermarks to be embedded. This choice conforms to the near-to-optimum performance of both QIM and SCS in the single-user case. However, we show that these schemes should be appropriately designed when it comes to the multiuser case. A joint design is required so as to closely approach the theoretical performance limits. For instance, for both the resulting BC- and MAC-based schemes, the improvement brought by this joint design is pointed out through comparison with the straightforward—rather intuitive—corresponding scheme which is obtained by simply superimposing (i.e., with no joint design) scalar schemes (or DPCs for the ideal coding). We introduce the notion of "awareness" to refer to this joint design. An interesting contribution at this stage is then that awareness helps in improving system performance. Awareness in the BC case basically implies that the encoder responsible for embedding the robust watermark is aware that a fragile signal is also embedded (with a known power), and thus, it modifies the coding scheme accordingly. This allows increasing the rate for the robust watermark. Similarly, awareness in the MAC case takes advantage at the embedder from the knowledge that a peeling-off decoder is used, i.e., that the better watermark is subtracted, an operation that changes the channel seen by the embedder. Again, the way to account for this MAC-awareness is to change the coding parameters. This increases the rate at which the worse watermark can be reliably communicated. The improvement brought up by awareness is demonstrated through both achievable rate region and bit error rate (BER) analysis. We finally show that performance can further be made closer to the theoretical limits by considering lattice-based codebooks. Some finite-dimensional lattices with good packing and quantization properties are considered for illustration.

[1]The materials in this paper have been partially published in [1] and [18].

The rest of this paper is organized as follows. After introducing the notation, we recall in Section II some fundamental principles of the dirty paper coding (DPC) technique. Also, we give a brief review of the formal statement of the information embedding problem as communication with side information available only at the transmitter, together with the state-of-the-art suboptimal practical coding schemes. These schemes will serve as baseline for the construction of the proposed approaches throughout this paper. Then, we turn in Section III to a detailed discussion on multiple information embedding applications. Two mathematical models corresponding to the multiple information embedding problem viewed either as communication over a degraded BC with state information at the transmitter or as communication over an MAC with state information at the transmitters are provided. Corresponding performance analyses are undertaken in Sections IV-A and IV-B, respectively. For each of these two mathematical models, analysis is carried out within the context of two watermarks using scalar-valued codebooks. Section V extends these results to the more general case of an arbitrary number of watermarks using high dimensional lattice-based codebooks. Finally, we close with a discussion followed by some concluding remarks in Section VI.

A. Notation

Throughout the paper, boldface fonts denote vectors. We use uppercase letters to denote random variables, lowercase letters for their individual values, e.g., $\mathbf{x} = (x_1, x_2, \ldots, x_N)$, and calligraphic fonts for sets, e.g., $\mathcal{X}$. Unless otherwise specified, vectors are assumed to be in the n-dimensional Euclidean space $(\mathbb{R}^n, \|\cdot\|)$ where $\|\cdot\|$ denotes the Euclidean norm of vectors. For a generic random vector $\mathbf{X}$, we use $\mathbb{E}_{\mathbf{X}}[\cdot]$ to denote the expectation taken with respect to $\mathbf{X}$ and $f_{\mathbf{X}}(.)$ to denote its probability density function (pdf). The Gaussian distribution with mean μ and square deviation σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$. A random variable $\mathbf{X}$ with conditional pdf given $\mathbf{S}$ is denoted by $\mathbf{X}\,|\,\mathbf{S}$.

II. Information Embedding and DPC

In this section, we first give a brief review of the information embedding problem as DPC. The resulting framework uses DPC principles to provide the ultimate theoretical performance

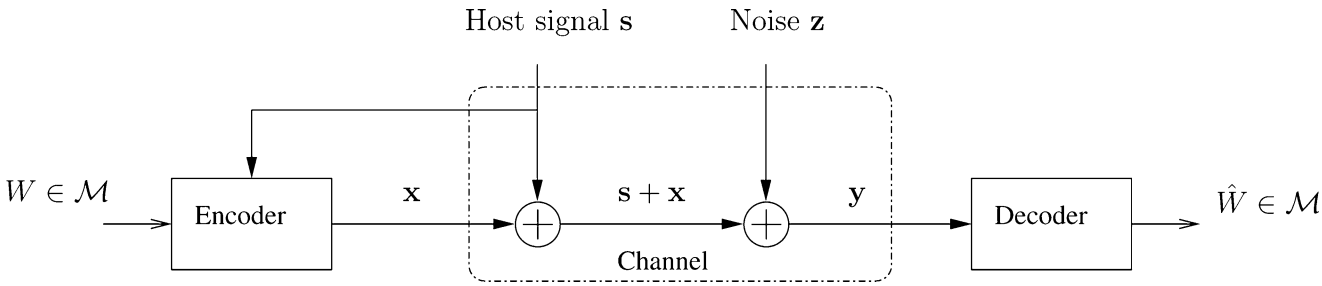


Fig. 1. Blind information embedding viewed as DPC over a Gaussian channel.

which is used as baseline for comparison in the rest of this paper. Next, both the well-known SCS [5] and QIM [3] are briefly reviewed together with their achievable performance.

A. Information Embedding as Communication With Side Information

Fig. 1 depicts a block diagram of the blind information embedding problem considered as a communication problem. A message m has to be sent to a receiver through some channel called the *watermark channel*. This channel is assumed to be i.i.d. Gaussian. We denote the Gaussian channel noise by $\mathbf{Z}$, with $Z_i \sim \mathcal{N}(0,N)$. The message m may be represented by a sequence $\{W\}$ of $\mathcal{M}$-ary symbols, with $\mathcal{M}=\{1,\ldots,M\}$, so as the transmission of the message m amounts to that of the corresponding symbols $\{W\}$. Thus, from now on, we will concentrate on the reliable transmission of W. Also, we will loosely use the term "message" to refer to the symbol W itself, instead of m. Prior to transmission, the message W is encoded into a signal $\mathbf{X}$ called the watermark which is then embedded into the cover signal $\mathbf{S} \in \mathbb{R}^n$, thus forming the watermarked or composite signal $\mathbf{S}+\mathbf{X}$. We assume that the cover signal $S_i \sim \mathcal{N}(0,Q)$ is Gaussian i.i.d. distributed and the watermarker $\mathbf{X}$ must satisfy the input power constraint $\mathbb{E}[\mathbf{X}^2] \leq P$. M is the greatest integer smaller than or equal to 2^{nR} and R is the transmission rate, expressed in number of bits per host sample that the encoder can reliably transmit. The watermark must be embedded without introducing any perceptible distortion to the host signal. This corresponds to the input power constraint in conventional power-limited communication and is commonly called the *transparency* requirement. The *robustness* requirement refers to the ability of the watermark to survive channel degradations. Rather than considering watermarking as communication over a very noisy channel where the cover signal $\mathbf{S}$ acts as self-interference as in SSM, it has been realized [20], [4] that blind watermarking can be viewed as communication with state information noncausally known at the transmitter, the state information being the cover signal $\mathbf{S}$ (entirely known at the transmitter). The relevant work is the initial Costa's "Writing on Dirty Paper" [6], also commonly known as DPC. Costa was the first to show the remarkable result that the interference $\mathbf{S}$, noncausally known only to the encoder, incurs no loss in capacity relative to the standard interference-free additive white Gaussian noise (AWGN) channel, i.e.,

$$C = \frac{1}{2}\log\left(1+\frac{P}{N}\right). \tag{1}$$

The achievability of this capacity of dirty paper channels is based on random binning arguments for general channels with state information [7]. This is based on random construction of a Gaussian codebook $\{\mathcal{U}_1,\ldots,\mathcal{U}_M\}$ and random partitioning of its codewords into "bins." In the Gaussian case (side information $\mathbf{S}$ and noise $\mathbf{Z}$ i.i.d. Gaussian), Costa showed that with the choice of the input distribution $p(u,x|s)$ such that

$$\mathbf{X} \sim \mathcal{N}(0,P) \quad \text{independent of } \mathbf{S} \tag{2a}$$
$$\mathbf{U} = \mathbf{X} + \alpha\mathbf{S} \quad \text{with } \alpha = P/(P+N) \tag{2b}$$

the capacity (1) is attained. This ideal DPC is, however, not feasible in practice due to the huge random codewords size needed for efficient binning. Therefore, some suboptimal lower complexity practical schemes have been proposed in [3] and [5]. A brief review is given in Section II-B.

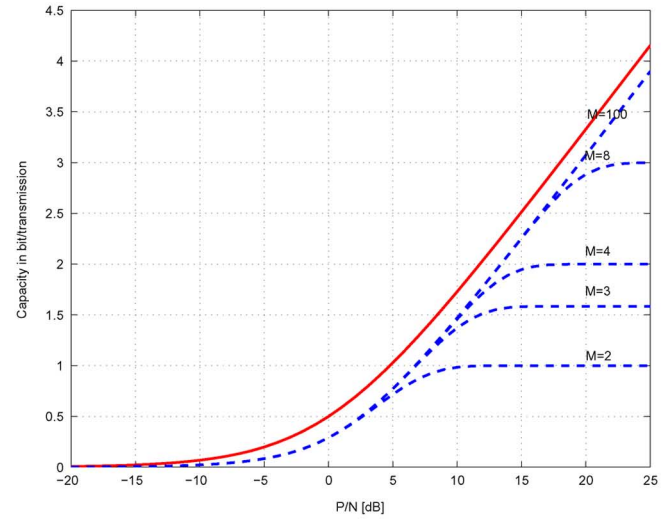
B. Suboptimal Coding

Following Costa's ideal DPC, Chen *et al.* proposed the use of structured quantization-based codebooks in [3]. The resulting embedding scheme is referred to as QIM. Whereas in [5], Eggers *et al.* designed a practical SCS where the random codebook $\mathbf{U}$ is chosen to be a concatenation of dithered scalar uniform quantizers. The watermark signal is a scaled version of the quantization error, i.e.,

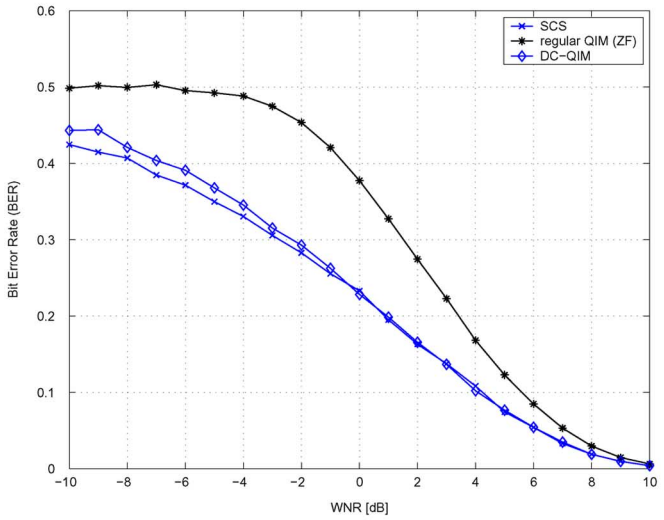
$$x_k = \tilde{\alpha}\left(\mathcal{Q}_\Delta\left(s_k - \frac{W}{M}\Delta\right) - \left(s_k - \frac{W}{M}\Delta\right)\right) \tag{3}$$

with $\Delta = \sqrt{12P}/\tilde{\alpha}$, $\tilde{\alpha} = \sqrt{P/(P+2.71N)}$ and $\mathcal{Q}_\Delta$ is the uniform scalar quantizer with constant step size Δ. Decoding is also based on scalar quantization of the received signal $\mathbf{y} = \mathbf{x}+\mathbf{s}+\mathbf{z}$ followed by a thresholding procedure. That is, the estimate $\widehat{W}$ of the transmitted message W is the closest integer to $r_k M/\Delta$, with $r_k = \mathcal{Q}_\Delta(y_k) - y_k$. The optimum parameter $\tilde{\alpha} = \sqrt{P/(P+2.71N)}$ is obtained by numerically maximizing the Shannon mutual information $I(W;r)$.[2] With this setting, SCS performs close to the optimal DPC. The aforementioned QIM which corresponds to the inflation parameter $\alpha = 1$ is less efficient, especially at relatively high noise levels. This QIM embedding function is referred to as *regular* QIM. Regular QIM can be slightly modified so as to increase its immunity to noise. The resulting scheme, called distortion-compensated QIM (DC-QIM), corresponds to $\alpha = P/(P+N)$ and performs very close to SCS, as shown in Fig. 2. We observe that SCS and DC-QIM

[2]Caution should be exercised here as $\mathbf{r}$ is the error quantization of the received signal, not the received signal itself.



(a)



(b)

Fig. 2. Performance of SCS, regular and DC-QIM in terms of the following: (a) capacity in bit per transmission and (b) BER. (a) M-ary SCS capacity (dashed) and full AWGN capacity (solid). (b) SCS outperforms—by far—regular QIM in terms of BER. A slight improvement over DC-QIM is observed at very low watermark-to-noise ratio (WNR) $= 10\log_{10}(P/N)$.

schemes, though clearly suboptimal, perform close to the ideal DPC. This constitutes the main motivation focus adapting them to the multiple watermarking situation.

III. Multiple User Information Embedding: Broadcast and MAC Setups

In an information embedding context, "multiple user" refers to the situation where several messages W_i have to be embedded into a common cover signal $\mathbf{S}$. The embedding may or may not require different robustness and transparency requirements. This means that each of these messages can be *robust*, *semifragile*, or *fragile*. Also, depending on the targeted application, the watermarking system may require either joint or separate decoding. For joint decoding, think of one single *trusted* authority checking for several (say K) watermarks at once. For separate (or distributed) decoding, think of several (say L) authorities each checking for its own watermark. In order to emphasize the very general case, one may even imagine these decoders having access to different noisy versions of the same watermarked content. This is due to the possibly different channel degradations the watermarked content may experience depending on the receiver location (think of a watermarked image being transmitted over a mobile network, with watermarking verification performed at different nodes of this network). As in decoding process, we may wish that the encoding of these messages be performed either jointly or separately. Some of the situations of concern are given by the illustrative examples described previously, with the receivers playing the role of the transmitters and vice-versa. Of course, though intentionally kept in its very general form, this model may not include some specific multiple information embedding situations. This is due to the difficulty of finding a single unifying approach. Nevertheless, the framework that we proposed is sufficiently general to involve the most important multiple information embedding scenarios. For instance, two classes of such scenarios, that we will recognize as being equivalent to communication over a degraded BC and an MAC in Sections III-A and III-B, respectively, are worthy of deep investigations. To simplify the exposition, we first restrict our attention to the two-watermarks embedding scenario. Then, extension to the general case follows.

A. Mathematical Model for BC-Like Multiuser Information Embedding

Consider an information embedding system aiming at embedding two messages W_1 and W_2, assumed to be M_1-ary and M_2-ary, respectively, into the same cover signal $\mathbf{S} \sim \mathcal{N}(0, Q)$. We suppose that one single *trusted authority* (the same encoder) has to embed these two messages and that embedding should be performed in such a way that the corresponding two watermarks correspond to two different usages (separate decoders). For example, the watermark $\mathbf{X}_2$ (carrying W_2) should be very robust whereas the watermark $\mathbf{X}_1$ (carrying W_1) may be of lesser robustness. This means that the watermark $\mathbf{X}_2$ must survive channel degradations up to some noise level N_2 larger than N_1, i.e., $N_2 \gg N_1$. Furthermore, the previously mentioned transparency requirement implies that the two watermarks put together must satisfy the input power constraint P, i.e., $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2$ is constrained to have $\mathbb{E}_{\mathbf{X}}[\mathbf{X}^2] = P$. Assuming independent watermarks[3] $\mathbf{X}_1$ and $\mathbf{X}_2$, we suppose with no loss of generality that $\mathbb{E}_{\mathbf{X}_1}[\mathbf{X}_1^2] = \gamma P$ and $\mathbb{E}_{\mathbf{X}_2}[\mathbf{X}_2^2] = (1-\gamma)P$, where $\gamma \in [0,1]$ may be arbitrarily chosen to share power between both watermarks.

In practice, this multiple watermarking scenario can be used to serve multiple purposes. In the scope of watermarking of medical images, for example, we may wish to store the patient information into the corresponding image, in a secure and private way. This information is sometimes called the "annotation part" of the watermark and is hence required to be sufficiently robust. Further, we may wish to use an additional possibly fragile "tamper detection part" to detect tampering. Another example stems from proof-of-ownership applications: We

[3]A justification of this assumption will be provided in Section IV.

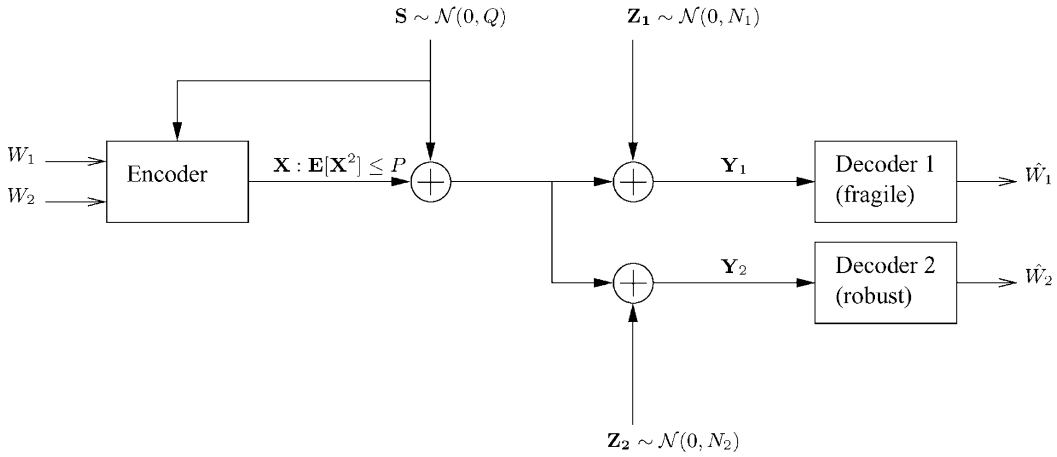


Fig. 3. Two users information embedding viewed as communication over a two-users GBC.

may wish to use one watermark to convey ownership information (should be robust) and a second watermark to check for content integrity (should be semifragile or fragile). A third example concerns watermarking for distributed storage. Suppose that a multimedia content (e.g., video or audio) has to be stored in different storage devices. Furthermore, we want to protect this multimedia content against piracy, by the use of a watermark. As the alteration level induced by the storage and extraction processes may differ from one device to another, the encoding technique must enable the reliably decoded rate to adapt to the actual alteration level. Of course many other examples and applications can be listed. We just mention here that the model at hand can be applied every time one watermarking authority (i.e., one transmitter) has to simultaneously embed several watermarks in such a way that these watermarks satisfy different robustness requirements.

Assuming Gaussian channel noises $\mathbf{Z}_i \sim \mathcal{N}(0, N_i)$, with $i = 1, 2$, a simplified block diagram of the transmission scheme of interest is shown in Fig. 3. Decoder i decodes $\widehat{W}_i$ from the received signal $\mathbf{Y}_i = \mathbf{X}_1 + \mathbf{X}_2 + \mathbf{S} + \mathbf{Z}_i$ at rate R_i. An error occurs if $\widehat{W}_i \neq W_i$. Functionally, this is the very transmission diagram of a two users Gaussian broadcast channel (GBC) with state information available at the transmitter but not at the receivers. In addition, the watermark $\mathbf{X}_2$ having to be robust plays the role of the message directed to the "degraded user" in a broadcast context. Conversely, the watermark $\mathbf{X}_1$ plays the role of the message directed to the "better user." Also, here, we have considered only two watermarks. The similarity with an L-users BC will be retained if, instead of just two watermarks, L watermarks are to be simultaneously embedded by the same so-called *trusted* authority.

B. Mathematical Model for MAC-Like Multiuser Information Embedding

We now consider another situation. Again, the watermarking system aims at embedding two independent messages W_1 and W_2 into the same cover signal $\mathbf{S}$. However, the present situation is different in that, this time, 1) embedding is performed by two different authorities, each having to embed its own message satisfying a given power requirement and 2) at the receiver, a single *trusted* authority checks for both watermarks. We assume no particular cooperation between the two embedding authorities, meaning that the watermarks $\mathbf{X}_1$ (carrying W_1) and $\mathbf{X}_2$ (carrying W_2) should be designed independently of each other and should satisfy independent power constraints $\mathbb{E}[\mathbf{X}_i^2] \leq P_i, i = 1, 2$. Note that, in addition, the composite watermark $\mathbf{X}_1 + \mathbf{X}_2$ with power at most equal $P = P_1 + P_2$ must satisfy the fidelity criterion to the nonwatermarked content. However, the power constraint here is fundamentally different from that in the aforementioned BC setup, since individual power constraints must be satisfied independently.

In practice, this multiple watermarking scenario can be used to serve multiple purposes. Loosely speaking, every watermarking system addressing the same application multiple times is concerned. An example stemming from proof-of-ownership applications is as follows. Consider two different creators independently watermarking the same original content $\mathbf{S}$, as it is common for large artistic works such as feature films and music recordings. Each of the two watermarks may contain private information. A common *trusted* authority may have to check for both watermarks. This is the case when an authenticator agent needs to track down the initial owner of an illegally distributed image, for example. A second example is the so-called hybrid in-band on-channel digital audio broadcasting [3]. In this application, we would like to simultaneously transmit two digital signals within the same existing analog (AM and/or FM) commercial broadcast radio without interfering with conventional analog reception. Thus, the analog signal is the cover signal and the two digital signals are the two watermarks. These two digital signals may be designed independently. One digital signal may be used as an enhancement to refine the analog signal and the other as supplemental information such as station or program identification. A third application concerns distributed (i.e., at different places) watermarking: Some fingerprinting can be embedded right at the camera, while possible annotations can be added next to the storage device.

Assuming a Gaussian channel noise $\mathbf{Z} \sim \mathcal{N}(0, N)$ corrupting the watermarked signal $\mathbf{S} + \mathbf{X}$, a simplified diagram is shown in Fig. 4. The encoder $i, i = 1, 2$, encodes W_i into $\mathbf{X}_i$ at rate R_i. The decoder outputs $(\widehat{W}_1, \widehat{W}_2)$. An error occurs if

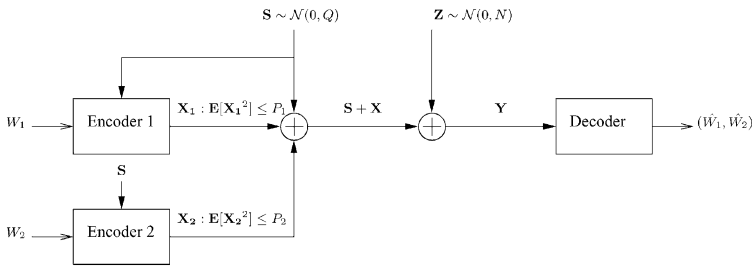


Fig. 4. Two users information embedding viewed as communication over a (two users) MAC.

$(\widehat{W_1}, \widehat{W_2}) \neq (W_1, W_2)$. Functionally, this is the very transmission diagram of a two users Gaussian MAC (GMAC) with state information available at the transmitters but not at the receiver. Note that, here, we have considered only two watermarks. The similarity with a K-users MAC will be retained if, instead of just two authorities, K different embedding authorities, each encoding its own message, are considered.

The previous discussion indicates that there are strong similarities between multiple information embedding and conventional multiple user communication. In Sections IV and V, we rely on recent findings in multiuser information theory [19] to devise efficient implementable multiple watermarking schemes and address their practical achievable performance. Also, in our attempt to further highlight the analogy with conventional multiuser communication, we will sometimes use the terms "multiple users," "degraded user," and "better user" to loosely refer to "multiple watermarks," "the receiver decoding the more noisy watermarked content" and "the receiver decoding the less noisy watermarked content," respectively.

IV. Information Embedding Over Gaussian Broadcast and MACs

In this section, we are interested in designing efficient low-complexity multiuser information embedding schemes for each of the two situations considered in Section III. We first present a straightforward rather intuitive method based on superimposing two SCSs. This simple method can be thought as being "coding-unaware." Next, we use the similarity between multiuser information embedding problem and transmission over Gaussian BC and MAC to design more efficient multiple watermarking schemes. We refer to these latter strategies as being "broadcast-" and "MAC-aware," respectively. The improvement brought by "awareness" is illustrated through both achievable rate regions and BER enhancements. Note that we will assume, throughout this section, that the flat-host assumption is satisfied as long as quantization is concerned.

A. Broadcast-Aware Coding for Two-Users Information Embedding

A simple approach for designing a coding system for the two users information embedding problem considered in Section III-A consists in using two independent single-user DPCs (or SCSs for the corresponding suboptimal practical implementation).[4]

[4]Note that this is not the most naive design, each DPC being tuned based on all information available.

1) Broadcast-Unaware Coding (Double DPC): In essence, the ideal coding is based on successive encoding at the transmitter as follows.

1) Use a first DPC (denoted by DPC2) taking into account the known state $\mathbf{S}$ and the power of unknown noise $\mathbf{Z}_2$ to form the most robust watermark $\mathbf{X}_2$ intended to the degraded user. By using (2), DPC2 is given by $\mathbf{X}_2 = \mathbf{U}_2 - \alpha_2\mathbf{S}$ with
$$\mathbf{U}_2|\mathbf{S} \sim \mathcal{N}(\alpha_2\mathbf{S}, (1-\gamma)P), \qquad \text{with } \alpha_2 = \frac{(1-\gamma)P}{(1-\gamma)P+N_2}. \tag{4}$$
2) Use a second DPC (denoted by DPC1) taking into account the known state $\mathbf{S}+\mathbf{X}_2$, sum of the cover signal $\mathbf{S}$ and the already formed watermark $\mathbf{X}_2$, and the power of unknown noise $\mathbf{Z}_1$ to form the less robust watermark $\mathbf{X}_1$ intended to the better user. By using (2), DPC1 is given by $\mathbf{X}_1 = \mathbf{U}_1 - \alpha_1(\mathbf{S}+\mathbf{X}_2)$ with
$$\mathbf{U}_1|\mathbf{U}_2, \mathbf{S} \sim \mathcal{N}(\alpha_1(\mathbf{S}+\mathbf{X}_2), \gamma P), \qquad \text{with } \alpha_1 = \frac{\gamma P}{\gamma P+N_1}. \tag{5}$$
3) Finally, transmit the composite signal $\mathbf{S}+\mathbf{X}$ over the watermark channel, with $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2$ being the composite watermark. The received signals are $\mathbf{Y}_1 = \mathbf{X} + \mathbf{S} + \mathbf{Z}_1$ and $\mathbf{Y}_2 = \mathbf{X} + \mathbf{S} + \mathbf{Z}_2$.

Note that the watermark $\mathbf{X}_2$ should be embedded first because of the following intuitive reason. When considering the extreme case where the watermark $\mathbf{X}_1$ is fragile, this watermark should be, by design, damaged by any operation that alters the cover signal $\mathbf{S}$. Since robust embedding is such an operation, the fragile watermark should be embedded last. The theoretical achievable region $\mathcal{R}_{\mathrm{BC}}$ with DPC1 and DPC2 is given by
$$\mathcal{R}_{\mathrm{BC}}(P) = \bigcup_{0\leq\gamma\leq 1} \left\{ (R_1, R_2) : R_1 \leq \frac{1}{2}\log_2\left(1+\frac{\gamma P}{N_1}\right), \right. \\ \left. R_2 \leq R(\alpha_2, (1-\gamma)P, Q, \gamma P + N_2) \right\} \tag{6}$$
where $R(\alpha, P, Q, N) = (1/2)\log_2(P(P+Q+N)/(PQ(1-\alpha)^2 + N(P+\alpha^2 Q)))$ and Q is the power of the host signal $\mathbf{S}$. Using straightforward algebra, which is omitted for brevity, it can be shown that the rates in (6) can be obtained by evaluating the achievable region [19]
$$\mathcal{R}_{\mathrm{BC}}(P_{U_1U_2\,|\,S}) = \{(R_1, R_2) : \\ R_1 \leq I(U_1; Y_1\,|\,U_2) - I(U_1; S\,|\,U_2) \tag{7a}$$
$$R_2 \leq I(U_2; Y_2) - I(U_2; S)\} \tag{7b}$$
with the choice of $p(u_1, u_2, x\,|\,s)$ given by (5) and (4).

Using (3) and following the way a single-user SCS is derived from the corresponding single-user DPC, a suboptimal practical two-users scalar information embedding scheme can be derived by independently superimposing two SCSs (denoted by SCS1 and SCS2 and taken as scalar versions of DPC1 and DPC2, respectively). SCS1 and SCS2 are applied sequentially, starting with SCS2 for the design of the watermark $\mathbf{x}_2$ as an appropriate scaled version of the quantization error of the cover signal $\mathbf{s}$. Then, SCS1 designs the watermark $\mathbf{x}_1$ as an appropriate scaled version of the quantization error of the sum signal $\mathbf{s}+\mathbf{x}_2$. The corresponding uniform scalar quantizers $\mathcal{Q}_{\Delta_1}$ and $\mathcal{Q}_{\Delta_2}$ have step sizes $\Delta_1 = \sqrt{12\gamma P/\widetilde{\alpha_1}}$ and $\Delta_2 = \sqrt{12(1-\gamma)P/\widetilde{\alpha_2}}$, where

$$(\widetilde{\alpha_1},\widetilde{\alpha_2}) = \left(\sqrt{\frac{\gamma P}{\gamma P + 2.71N_1}}, \sqrt{\frac{(1-\gamma)P}{(1-\gamma)P+2.71N_2}}\right). \tag{8}$$

Note that the flat-host assumption on signals $\mathbf{s}$ and $\mathbf{s}+\mathbf{x}_2$ is assumed to hold as supposed previously. We denote by $(\widetilde{R_1},\widetilde{R_2})$ the transmission throughput achieved by this setup. This rate pair is computed numerically. Results are depicted in Fig. 5 and are compared to the theoretical rate pair $(R_1,R_2)\in\mathcal{R}_{\mathrm{BC}}$ given by (6), for two examples of channel parameters. The noise in first example, (i.e., the one such that $P/N_2 = 0$ dB) may model a channel attack which has the same power as the composite watermark $\mathbf{X}=\mathbf{X}_1+\mathbf{X}_2$. The performance of this first approach is worthy of the following brief discussion.

1) From (6), we see that DPC1—as given by (5)—is optimal. The achievable rate R_1 corresponds to that of a channel with not only no interfering cover signal $\mathbf{S}$, but also no interference signal $\mathbf{X}_2$. Thus, the message W_1 can be sent at its maximal rate, as if it were embedded alone. From "decoder 1" point of view, the channel from W_1 to $\mathbf{Y}_1$ is functionally equivalent to a single-user channel from W_1 to $\mathbf{Y}_1' = \mathbf{Y}_1 - \mathbf{U}_2 = \mathbf{X}_1 + (1-\alpha_2)\mathbf{S} + \mathbf{Z}_1$, having just $(1-\alpha_2)\mathbf{S}$ as state information, not $\mathbf{S}+\mathbf{X}_2$. Yet, it is not that $\mathbf{Y}_1$ is a single-user channel, but rather that the amount of reliably decodable information W_1 is exactly the same as if W_1 were transmitted alone over $\mathbf{Y}_1'$. Moreover, DPC2—as given by (4)—is not optimal. The reason is that the achievable rate R_2 in (6) is inferior to $(1/2)\log_2(1+(1-\gamma)P/(\gamma P+N_2))$. The latter rate is that of a watermark signal subject to the full interference penalty from both the cover signal $\mathbf{S}$ and the watermark $\mathbf{X}_1$.
2) SCS1 performs close to optimality. The scalar channel having a message W_1 as input and the quantization error as output is functionally equivalent to that from W_1 to $\mathbf{r}_1' = \mathcal{Q}_{\Delta_1}(\mathbf{y}_1') - \mathbf{y}_1'$, where $\mathbf{y}_1'$ is the single-user channel suffering only partly from the interference $\mathbf{X}_2$.[5] The practical transmission rate over this channel is given by the mutual information $I(W_1;r_1')$, the maximum of which (i.e., $\widetilde{R_1}$) is obtained with the choice (8) of $\widetilde{\alpha_1}$. However, being derived from DPC2—which itself is nonoptimal—SCS2 is obviously suboptimal. Consequently, the previous choice of parameter $\widetilde{\alpha_2}$ does not maximize the mutual information $I(W_2;r_2)$, with $\mathbf{r}_2 = \mathcal{Q}_{\Delta_2}(\mathbf{y}_2) - \mathbf{y}_2$.

[5]Note that in the equivalent channel $\mathbf{y}_1' = \mathbf{x}_1 + (1-\alpha_2)\mathbf{s} + \mathbf{z}_1$, the watermark $\mathbf{x}_1$ is formed as a scaled version of the quantization error of the channel state $(1-\alpha_2)\mathbf{s}$ and not $\mathbf{s}+\mathbf{x}_2$ as before.

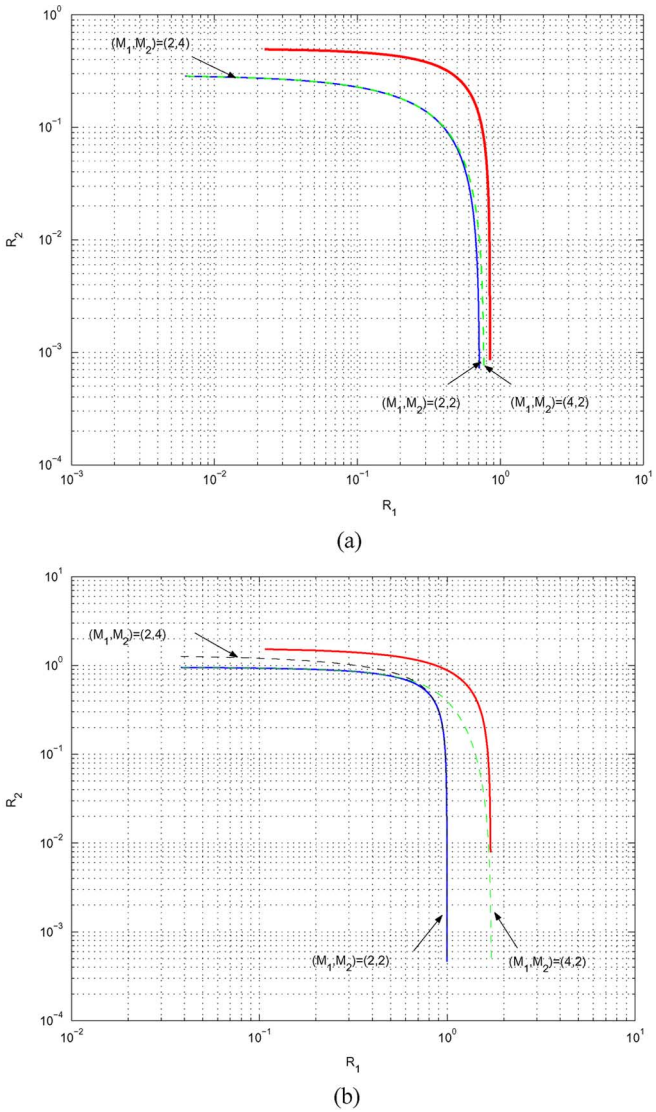


Fig. 5. Theoretical and feasible transmission rates for broadcast-unaware multiple user information embedding for two examples of SNR. For each SNR, the upper curve corresponds to the theoretical rate region $\mathcal{R}_{\mathrm{BC}}$ (6) of the double DPC and the lower curve corresponds to the achievable rate region $(\widetilde{R_1},\widetilde{R_2})$ of the two superimposed SCSs with quantization parameters given by (8). Dashed line correspond to (2-ary, 4-ary) and (4-ary, 2-ary) transmissions. (a) Rates for $P/N_1 = 5$ dB and $P/N_2 = 0$ dB. (b) Rates for $P/N_1 = 12$ dB and $P/N_2 = 9$ dB.

In Section IV-A2, we show that the encoding of W_2 can be improved so as to bring the rate $\widetilde{R_2}$ close to $R_2^{(\max)} = (1/2)\log_2(1+(1-\gamma)P/(\gamma P+N_2))$. The corresponding scheme, which we call "joint scalar DPC" in the sequel, improves system performance by making multiple information embedding broadcast-aware.

2) Broadcast-Aware Coding (Joint DPC): In Section III-A, we have shown that the communication scenario depicted in Fig. 3 is basically that of a degraded GBC with state information noncausally known to the transmitter but not to the receivers. In [19], it has been shown that the capacity region $\mathcal{C}_{\mathrm{BC}}$ of this

channel is given by

$$\mathcal{C}_{\mathrm{BC}}(P)=\bigcup_{0\le\gamma\le 1}\left\{(R_1,R_2): R_1\le\frac{1}{2}\log_2\left(1+\frac{\gamma P}{N_1}\right),\right.$$
$$\left. R_2\le\frac{1}{2}\log_2\left(1+\frac{(1-\gamma)P}{\gamma P+N_2}\right)\right\} \tag{9}$$

which is that of a GBC with no interfering signal $\mathbf{S}$. This region can be attained by an appropriate successive encoding scheme that uses two well-designed DPCs. The encoding of W_1 (DPC1) is still given by (5). For the encoding of W_2, however, the key point is to consider the unknown watermark $\mathbf{X}_1$ as noise. We refer to this by saying that the encoder is "aware" of the existence of the watermark $\mathbf{X}_1$ and takes it into account. The resulting DPC (again denoted by DPC2) uses the cover signal $\mathbf{S}$ as channel state and $\mathbf{Z}_2+\mathbf{X}_1$ as total channel noise

$$\mathbf{U}_2\,|\,\mathbf{S}\sim\mathcal{N}(\alpha_2\mathbf{S},(1-\gamma)P)\quad\text{with }\alpha_2$$
$$=\frac{(1-\gamma)P}{(1-\gamma)P+(N_2+\gamma P)} \tag{10a}$$
$$\mathbf{X}_2=\mathbf{U}_2-\alpha_2\mathbf{S}. \tag{10b}$$

Obviously, this encoding does not remove the interference due to $\mathbf{X}_1$. Nevertheless, DPC1 is optimal in that it attains the maximal possible rate $R_2^{(\max)}$ at which W_2 can be sent together with W_1.

3) Feasible Rate Region: Consider now a scalar implementation of this joint DPC scheme consisting in two successive SCSs. DPC2 can be implemented by a scalar scheme SCS2, quantizing the cover signal $\mathbf{s}$ and outputting the watermark $\mathbf{x}_2$ as an appropriate scaled version of the quantization error. We denote by $\widetilde{\alpha_2}$ and Δ_2 the corresponding scale factor and quantization step size, respectively. DPC1 can be implemented by a scalar scheme SCS1, quantizing the newly available signal $\mathbf{s}+\mathbf{x}_2$ and outputting the watermark $\mathbf{x}_1$ as an appropriately scaled version of the quantization error. We denote by $\widetilde{\alpha_1}$ and Δ_1 the corresponding scale factor and quantization step size, respectively. Let $\mathbf{Y}'_1=\mathbf{Y}_1-\mathbf{U}_2$ be the channel functionally equivalent to $\mathbf{Y}_1$ introduced previously. The resulting achievable rate region $\widetilde{\mathcal{R}}_{\mathrm{BC}}$, practically feasible with this coding, is given by

$$\widetilde{\mathcal{R}}_{\mathrm{BC}}(P)$$
$$=\bigcup_{0\le\gamma\le 1}\left\{(\widetilde{R_1},\widetilde{R_2}):\right.$$
$$\widetilde{R_1}\le\max_{\alpha_1\in[0,1]} I\left(W_1;\underbrace{\mathcal{Q}_{\Delta_1(\alpha_1,\gamma)}(\mathbf{y}'_1)-\mathbf{y}'_1}_{\mathbf{r}'_1}\right),$$
$$\left.\widetilde{R_2}\le\max_{\alpha_2\in[0,1]} I\left(W_2;\underbrace{\mathcal{Q}_{\Delta_2(\alpha_2,\gamma)}(\mathbf{y}_2)-\mathbf{y}_2}_{\mathbf{r}_2}\right)\right\}. \tag{11}$$

The proof simply follows from the previous discussion regarding the equivalent channels from W_1 to $\mathbf{r}'_1$ for the message W_1 and from W_2 to $\mathbf{r}_2$ for the message W_2. Each of these two channels conforms the single-user channel considered in [5] and has hence a similar expression of the transmission rate. The inflation parameters pair $(\widetilde{\alpha_1},\widetilde{\alpha_2})$ maximizing the right-hand side terms of (11) is given by

$$(\widetilde{\alpha_1},\widetilde{\alpha_2})=\left(\sqrt{\frac{\gamma P}{\gamma P+2.71N_1}},\right.$$
$$\left.\sqrt{\frac{(1-\gamma)P}{(1-\gamma)P+2.71(\gamma P+N_2)}}\right). \tag{12}$$

The region (11), obtained through a Monte-Carlo-based integration, is depicted in Fig. 6 and is compared to the ideal DPC region $\mathcal{C}_{\mathrm{BC}}$ given by (9), for two choices of channel parameters: weak channel noise [Fig. 6(c) and (d)] and strong channel noise [Fig. 6(a) and (b)]. The latter may model, for example, a channel attack with power equal to that of the composite watermark $\mathbf{X}=\mathbf{X}_1+\mathbf{X}_2$, as mentioned previously. Note that we need to compute the conditional probabilities $p_{\mathbf{r}'_1}(\mathbf{r}'_1\,|\,W_1)$ and $p_{\mathbf{r}_2}(\mathbf{r}_2\,|\,W_2)$. These are computed using the high-resolution quantization assumption $Q\gg P$, which is relevant in most watermarking applications. Improvement over the "double DPC" is made possible by increasing the rate R_2 at which the robust watermark can be sent. It is precisely "awareness" that allows such improvement. However, note that this improvement is more visible for high signal-to-noise ratio (SNR), as shown in Fig. 6(c). For low SNR, however, this improvement, though still theoretically possible as shown in Fig. 6(a), is very limited and is almost not visible for scalar codebooks. This can be interpreted as follows: The aforementioned "awareness," which can be viewed as a power saving technique for the "degraded user," does not sensibly improve the overall communication (i.e., increase transmission rate) when the channel is very bad.[6] Both theoretical and feasible rate regions of the BC-aware scheme are also depicted for nonbinary inputs in Fig. 6(d) and (b). It can be seen that, depending on the SNR, the practically feasible rate region (11) can more-or-less approach the theoretical capacity region $\mathcal{C}_{\mathrm{BC}}$ by increasing the sizes M_1 and M_2 of the input alphabets $\mathcal{M}_1$ and $\mathcal{M}_2$.[7]

4) BER Analysis and Discussion: Another performance analysis is based on measured BERs for hard-decision-based decoding of binary scalar DPC. Results are obtained with Monte-Carlo-based simulation and are depicted in Fig. 7. Note that the set of channel parameters chosen in Fig. 7 may model a wide range of admissible channel attacks on the individual watermarks, since the individual SNRs, $\mathrm{SNR}_1=10\log_{10}(\gamma P/N_1)$ and $\mathrm{SNR}_2=10\log_{10}((1-\gamma)P/(\gamma P+N_2))$, vary from -8 dB to 12 dB and from -15 dB to 9 dB, respectively, as the power-sharing parameter γ varies from 0 to unity. However, this may be not a good choice to model a strong attack on the composite watermark $\mathbf{X}_1+\mathbf{X}_2$ (for example, one such that $P/N_2=0$ dB). For such an attack, the individual rates are very low and the BERs are very bad. In principle, it would be possible to use any provably efficient error correction code for each of the channels $\mathbf{Y}_1$ and $\mathbf{Y}_2$ taken separately. However, at low SNR ranges,

[6]Note, however, that this should not be considered as a drawback since when the channel is very bad only little information is transmitted.

[7]However, a gap of about 1.53 dB should remain visible, i.e., $R_1-\widetilde{R_1}>1.53$ dB and $R_2-\widetilde{R_2}>1.53$ dB.

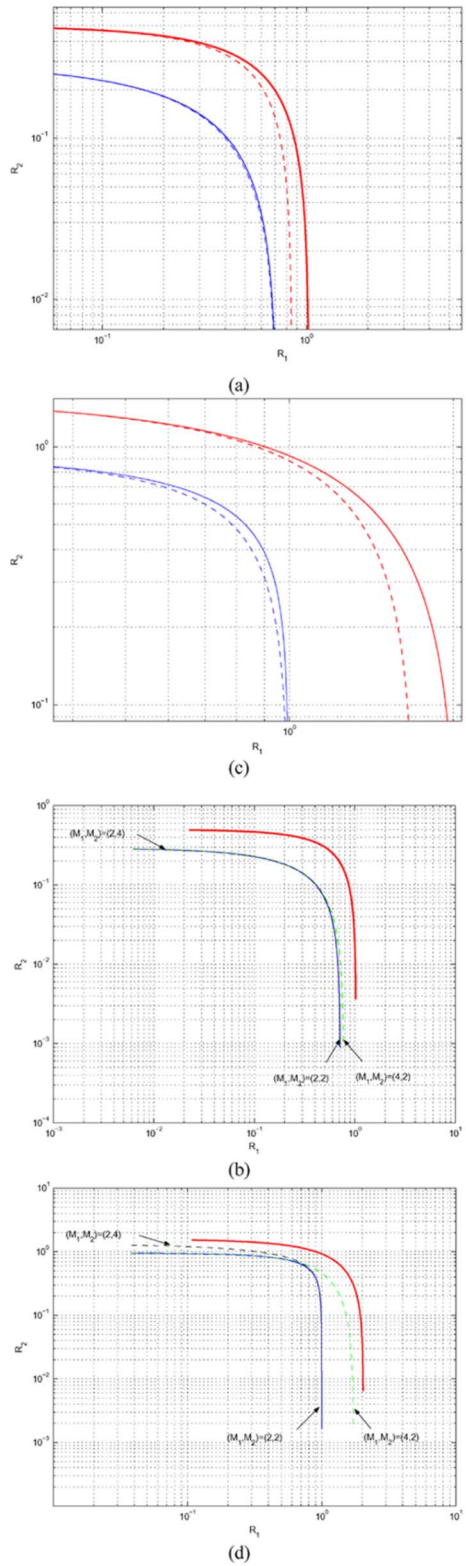


Fig. 6. Improvement brought by "BC-awareness" (with binary inputs) is depicted for (a) $P/N_1 = 5$ dB and $P/N_2 = 0$ dB and (c) $P/N_1 = 12$ dB and $P/N_2 = 9$ dB. Solid line corresponds to the rate region of the BC-aware scheme achievable theoretically (upper) and practically (lower). Dashed line corresponds to the rate region of the BC-unaware scheme achievable theoretically (upper) and practically (lower). (b) and (d): achievable rate region of the BC-aware scheme for M_1-ary and M_2-ary alphabets depicted for (b) $P/N_1 = 5$ dB and $P/N_2 = 0$ dB and (d) $P/N_1 = 12$ dB and $P/N_2 = 9$ dB.

it is well known that repetition coding is almost optimal. The curves in Fig. 7(a) are obtained with $(\rho_1, \rho_2) = (4, 4)$, meaning that W_1 and W_2 are repeated four times each. We observe that as $\gamma \in [0, 1]$ increases, the power part of the signal $\mathbf{X}$ allocated to the watermark carrying W_1 becomes larger and that allocated to the watermark carrying W_2 becomes smaller. This causes the corresponding BER curves to monotonously decrease and increase, respectively. Also, it can be checked that, when plotted separately, these curves are identical to those of an SCS with a signal-to-noise power ratio equal to SNR_1 and SNR_2, respectively. This conforms the assumption made previously regarding the functionally equivalent channels $\mathbf{y}_1'$ and $\mathbf{y}_2$. The curves depicted in Fig. 7 also motivate the following discussion.

1) In practical situations, the repetition factors ρ_1 and ρ_2 should be chosen in light of the desired transmission rates and robustness requirements. The choice $(\rho_1, \rho_2) = (4, 4)$ made previously should be taken just as a baseline example. Channel coding as a means of providing additional redundancy obviously strengthens the watermark immunity to channel degradations. However, such a redundancy inevitably limits the transmission rate. This means that for equal targeted transmissions rates R_1 and R_2, the repetition factors ρ_1 and ρ_2 should satisfy $\rho_2 \geq \rho_1$.
2) The scalar DPC considered here for multiple watermarking is constructed using insights from coding for broadcast channels [21], [22], as mentioned previously. Interestingly, in such channels, the user who experiences the better channel (less noisy) has to reliably decode the message assigned to the (degraded) user who experiences the worst channel (more noisy). In an information embedding context, this means that the robust watermark, which is supposed to survive channel degradation levels up to N_2, should be reliably decodable if, actually, the channel noise is less powerful. However, this strategy, which is inherently related to the principle of superposition coding at the transmitter combined with successive decoding (peeling-off technique) at the "better user" (decoder 1) [23], makes more sense in the situations where the "better user" is unable to reliably decode its own message if it does not primarily subtract off the interference due to the message assigned to the "degraded user." The DPC-based scheme is fundamentally different in that the interference is already subtracted off at the encoder. As a consequence, the "better user" does not need to decode the message of the degraded user.[8]
3) There could, however, be advantages and disadvantages for the DPC-based scheme described previously to follow such a strategy. An obvious disadvantage concerns security issues. In a transmission scheme where security is a major issue, the "better user" should not be able to reliably decode the message assigned to the "degraded user." By opposition, an obvious advantage stems from the following observation. If channel quality is improved, resulting in better SNR in the transmission of W_2, the "degraded user," being at present a "better user," should be able to reliably decode much more information W_2 than it does with the old channel quality. For the previously described DPC-based

[8]Note that by opposition to superposition coding, there is an important embedding ordering at the encoder. The benefit of such ordering is a decoupling of the receivers, and hence, a more scalable system. Each receiver needs only to know its own codebook to extract its message.

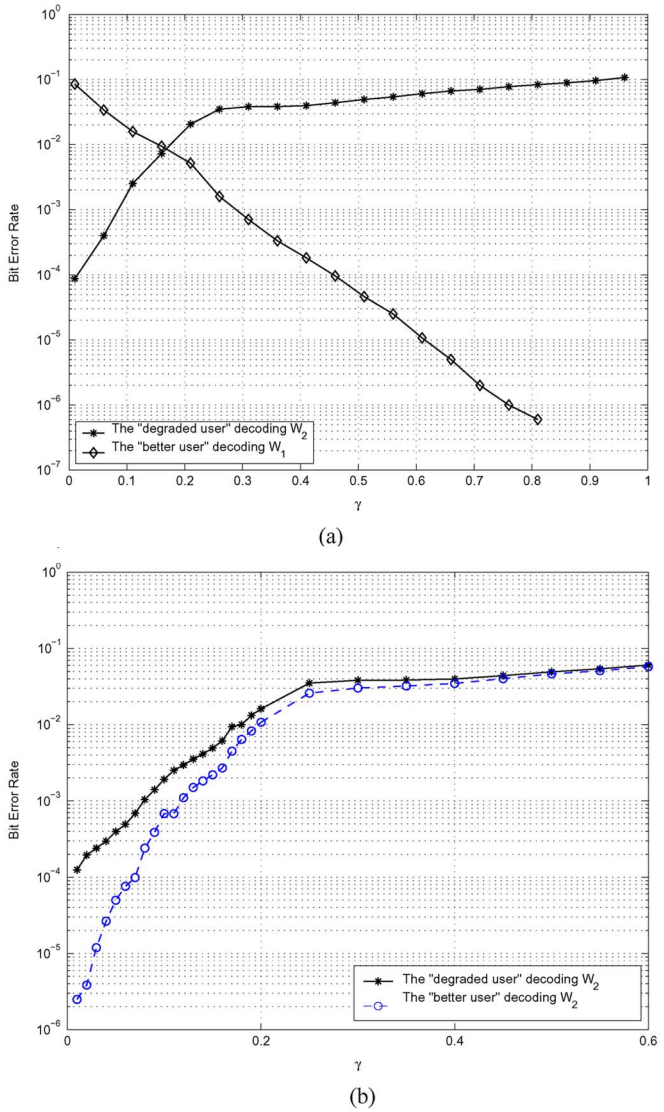


Fig. 7. Broadcast-aware multiple user information embedding. (a) BERs for binary transmission using repetition coding. (b) Each decoder can only decode "his" own watermark. Though much less noisy, the "better user" performs only slightly better than the "degraded user" in decoding message W_2. The messages W_1 and W_2 are repeated four times each, i.e., $(\rho_1, \rho_2) = (4, 4)$ and channel parameters are such that $P/N_1 = 12$ dB and $P/N_2 = 9$ dB.

scheme, to fulfill this additional requirement, one should focus on maximizing (over α_1) the conditional mutual information $I(W_1; r_1 | W_2)$. This would, however, lead to a suboptimal choice α_1' of the inflation parameter α_1 for the transmission of W_1, and consequently, to a smaller transmission rate $\widetilde{R_1} = I(W_1; r_1') \big|_{\alpha_1 = \widetilde{\alpha_1'}}$.

4) The present DPC scheme, as is, does not fully satisfy the aforementioned broadcast property. From Fig. 7(b), we observe that the "better user" does not fully exploit the fact of being much less noisy (than the degraded user) to more reliably decode W_2: The improvement in BER upon the "degraded user" is very small and is even negligible, as shown in Fig. 7(b). And even though this improvement seems to behave like the improvement in SNR (which is maximal at $\gamma = 0$), it is actually smaller than the one, $10\log_{10}((\gamma P + N_2)/(\gamma P + N_1))$ dB, which should be visible if the "better user" were able to reliably decode W_2 as in superposition coding.

B. MAC-Aware Coding for Two Users Information Embedding

In this section, we are interested in designing implementable multiple watermarking schemes for the situation described in Section III-B. Paralleling the development made in Section IV, we provide a performance analysis for two MAC-aware and unaware multiple watermarking strategies.

1) MAC-Unaware Coding (Double DPC): The situation described in Section III-B corresponds in essence to two "dirty paper" channels. A simple approach for designing a watermark system for this situation consists in two single-user DPCs (or SCSs for the corresponding practical implementation). Let $\mathbf{Y} = \mathbf{X}_1 + \mathbf{X}_2 + \mathbf{S} + \mathbf{Z}$ denote the received signal. Upon reception, the receiver should reliably decode the messages W_1 and W_2 having been embedded into the watermarks $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively. However, since decoding is performed jointly, the successful decoding of one of the two messages should benefit of the other message. This is illustrated through the following possible coding.

1) Encoder 2 uses a DPC (DPC2) taking into account the known state $\mathbf{S}$ and the power of unknown noise $\mathbf{Z}$ to form the watermark $\mathbf{X}_2$ of power P_2 and carrying W_2 as $\mathbf{X}_2 = \mathbf{U}_2 - \alpha_2\mathbf{S}$, where

$$\mathbf{U}_2 \sim \mathcal{N}(\alpha_2\mathbf{S}, P_2) \quad \text{with} \quad \alpha_2 = \frac{P_2}{P_2 + N}. \tag{13}$$

At reception, the decoder first decodes W_2, and then, cleans up the channel by subtracting the interference penalty $\mathbf{U}_2$ that the transmission of W_2 causes to that of W_1.[9] Thus, the channel for W_1 is made equivalent to $\mathbf{Y}_1 = \mathbf{Y} - \mathbf{U}_2 = \mathbf{X}_1 + (1 - \alpha_2)\mathbf{S} + \mathbf{Z}$. This "cleaning up" step is inherently associated with *successive* decoding and is sometimes referred to as the *peeling-off* technique. Hence, encoder 1 can reliably transmit W_1 over the channel $\mathbf{Y}_1$ by using a second DPC (DPC1).

2) Encoder 1 forms $\mathbf{X}_1$ as $\mathbf{X}_1 = \mathbf{U}_1 - \alpha_1\mathbf{S}$, where

$$\mathbf{U}_1 \,|\, \mathbf{S} \sim \mathcal{N}(\alpha_1\mathbf{S}, P_1) \quad \text{with} \quad \alpha_1 = (1 - \alpha_2)\frac{P_1}{P_1 + N}. \tag{14}$$

The rate pairs $(R_1, R_2) \in \mathcal{R}_{\text{MAC}}$ achieved by the considered two DPCs are those corresponding to the corner point (B1) of the achievable region $\mathcal{R}_{\text{MAC}}$ depicted in Fig. 8, and are given by

$$R_1(B1) = \frac{1}{2}\log_2\left(1 + \frac{P_1}{N}\right) \tag{15a}$$

$$R_2(B1) = \frac{1}{2}\log_2 \times \left(\frac{P_2(P_2 + Q + N + P_1)}{P_2 Q(1 - \alpha_2)^2 + (N + P_1)(P_2 + \alpha_2^2 Q)}\right). \tag{15b}$$

[9]Note that, theoretically, the decoder looks for the (unique) codeword $\mathbf{U}_2$ such that $(\mathbf{U}_2, \mathbf{Y})$ is jointly typical. In practice, however, the decoder only knows an estimate $\hat{\mathbf{U}}_2$ of the codeword $\mathbf{U}_2$ even if W_2 is decoded perfectly, since the host $\mathbf{S}$ is unknown at the receiver (see discussion in Section IV-B4).

Using straightforward algebra which is omitted for brevity, it can be shown that the rates in (15) correspond to a corner point in the rate region obtained by evaluating the achievable region [19]

$$\begin{aligned}\mathcal{R}_{\mathrm{MAC}}(P_{U_1U_2|S}) &= \{(R_1,R_2): \\ &\quad R_1 \le I(U_1;Y|U_2) - I(U_1;S|U_2) \\ &\quad R_2 \le I(U_2;Y|U_1) - I(U_2;S|U_1), \\ &\quad R_1+R_2 \le I(U_1,U_2;Y) - I(U_1,U_2;S),\} \quad (16)\end{aligned}$$

with the choice of $p(u_1,x_1,u_2,x_2|s) = p(u_1,x_1|s)\,p(u_2,x_2|s)$ given by (13) and (14). Following the same principle, similar DPC schemes allowing to attain the corner points (A), (C1), and (D) can be designed. The corner point (A) corresponds to the watermark $\mathbf{X}_1$ (i.e., the information W_1) being sent at its maximum achievable rate whereas the watermark $\mathbf{X}_2$ (i.e., the information W_2) not transmitted at all. The two corner points (C1) and (D) correspond to the points (B1) and (A), respectively, with the roles of the watermarks $\mathbf{X}_1$ and $\mathbf{X}_2$ reversed. Any rate pair lying on the lines connecting these corner points can be attained by time sharing. We concentrate on the corner point (B1) and consider a practical implementation of this theoretical setup. This can be performed by using two SCSs, SCS1 and SCS2, consisting of scalar versions of DPC1 and DPC2. The uniform scalar quantizers $\mathcal{Q}_{\Delta_1}$ and $\mathcal{Q}_{\Delta_2}$ have step sizes $\Delta_1 = \sqrt{12P_1}/\widetilde{\alpha_1}$ and $\Delta_2 = \sqrt{12P_2}/\widetilde{\alpha_2}$, where

$$(\widetilde{\alpha_1},\widetilde{\alpha_2}) = \left((1-\alpha_2)\sqrt{\frac{P_1}{P_1+2.71N}}, \sqrt{\frac{P_2}{P_2+2.71N}}\right) \quad (17)$$

conform the codebooks choice in (13) and (14).[10] Note that the signal $\mathbf{S}$ is assumed to be flat-host as mentioned previously. The feasible transmission rate pair achieved by this practical coding corresponds to the corner point (B1') in the diagrams shown in Fig. 8. Results are depicted for two choices of channel parameters: strong channel noise [shown in Fig. 8(a)] and weak channel noise [shown in Fig. 8(b)]. The strong noise may model a channel attack which has the same power as the composite watermark $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2$. The performance of this first approach can be summarized as follows.

1) From (15b), we see that DPC1—as given by (14)—is optimal. The interference due to the cover signal $\mathbf{S}$ and the second watermark $\mathbf{X}_2$ is completely canceled. Hence, the watermark $\mathbf{X}_1$ can be sent at its maximal rate R_1, as if it were alone over the watermark channel. The channel from W_1 to $\mathbf{Y}$ is functionally equivalent to that from W_1 to $\mathbf{Y}_1 = \mathbf{Y} - \mathbf{U}_2$. However, DPC2—as given by (13)—is nonoptimal, because the rate R_2 given by (15b) is inferior to $(1/2)\log_2(1+P_2/(P_1+N))$, which is that of a watermark subject to the full interference penalty from both the cover signal $\mathbf{S}$ and the watermark $\mathbf{X}_1$.
2) SCS1 performs close to optimality. The scalar channel is equivalent to that from W_1 to $\mathbf{r}_1 = \mathcal{Q}_{\Delta_1}(\mathbf{y}_1) - \mathbf{y}_1$. The practical transmission rate over this channel is given by the mutual information $I(W_1;r_1)$, the maximum of which (i.e., $\widetilde{R_1}$) is obtained with the choice (17) of $\widetilde{\alpha_1}$. However, SCS2 is nonoptimal, simply because DPC2 is not. The inflation parameter $\widetilde{\alpha_2}$ does not maximize the mutual information $I(W_2;r)$, with $\mathbf{r} = \mathcal{Q}_{\Delta_2}(\mathbf{y}) - \mathbf{y}$. Thus, the achievable rate $\widetilde{R_2}$ is not maximal and corresponds to $\widetilde{R_2} = I(W_2;r)|_{\alpha_2=\widetilde{\alpha_2}}$.

[10]Note that the choice $(\widetilde{\alpha_1},\widetilde{\alpha_2})$ in (17) does not maximize the input–output mutual information. Rather, it directly traces the way in which the codebooks are generated in (13) and (14).

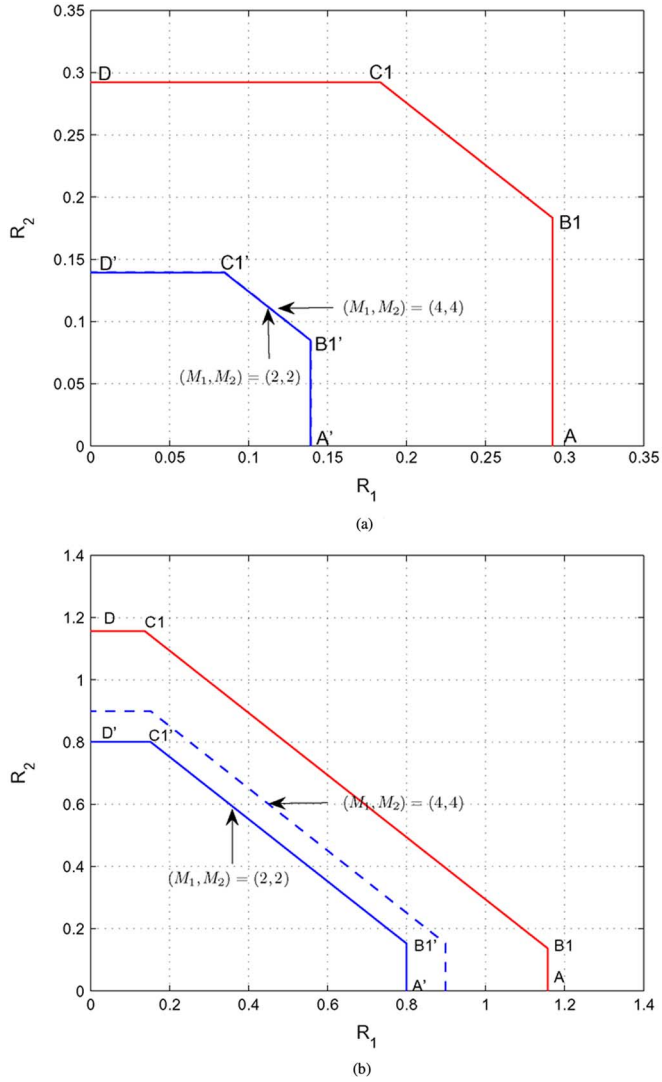


Fig. 8. Theoretical and feasible transmission rates for MAC-unaware multiple user information embedding. The frontier with corner points (A), (B1), (C1), and (D) corresponds to the theoretical rate pair $(R_1,R_2) \in \mathcal{R}_{\mathrm{MAC}}$ of the double ideal DPC. The frontier with corner points (A'), (B1'), (C1'), and (D') corresponds to the feasible rate pair $(\widetilde{R_1},\widetilde{R_2})$ of the two superimposed SCSs. Dashed line corresponds to practical rates obtained with the use of quaternary alphabets. (a) Rates for $P_1 = P_2; (P_1+P_2)/N = 0$ dB. (b) Rates for $P_1 = P_2; (P_1+P_2)/N = 9$ dB.

The encoding of W_2 can be improved so as to bring the achievable rate $\widetilde{R_2}(B1')$ close to $R_2^{(\max)} = (1/2)\log_2(1+(P_2)/(P_1+N))$. The corresponding scheme, called "joint DPC," enhances the performance by making multiuser information embedding MAC-aware.

2) MAC-Aware Coding (Joint DPC): In Section III-B, we argued that the communication scenario depicted in Fig. 4 is basically that of a GMAC with state information noncausally known to the transmitters but not to the receiver. In [19], it is

reported that the capacity region $\mathcal{C}_{\mathrm{MAC}}$ of this channel is given by

$$\mathcal{C}_{\mathrm{MAC}}(P_1,P_2)=\Bigg\{(R_1,R_2):\\ R_1\le\frac{1}{2}\log_2\left(1+\frac{P_1}{N}\right),\\ R_2\le\frac{1}{2}\log_2\left(1+\frac{P_2}{N}\right),\\ R_1+R_2\le\frac{1}{2}\log_2\left(1+\frac{P_1+P_2}{N}\right)\Bigg\} \tag{18}$$

which is that of a GMAC with no interfering signal $\mathbf{S}$. This region, with corner points (A), (B), (C), and (D), is shown in Fig. 9 and can be attained by an appropriate successive encoding scheme that uses well-designed DPCs. Consider, for example, the corner point (B). The encoding of W_1 is again given by (14), recognized previously to be optimal.[11] The encoding DPC2 of W_2, however, should be changed so as to consider the watermark $\mathbf{X}_1$ as noise. We refer to this situation by saying that the encoder should be "aware" of the existence of $\mathbf{X}_1$ and acts accordingly. The resulting DPC (again denoted by DPC2) uses the cover signal $\mathbf{S}$ as channel state and the signal $\mathbf{Z}+\mathbf{X}_1$ as total channel noise

$$\mathbf{U}_2\,|\,\mathbf{S}\sim\mathcal{N}(\alpha_2\mathbf{S},P_2)\quad\text{with}\quad\alpha_2=\frac{P_2}{P_2+(P_1+N)}. \tag{19}$$

Obviously, the interference due to $\mathbf{X}_1$ is not removed. However, this scheme is optimal in that it achieves the maximum rate $R_2^{(\max)}$ at which the message W_2 can be sent as long as the message W_1 is sent at its maximum rate.

3) Feasible Rate Region: We consider now a practical implementation for this joint scheme through two jointly designed SCSs with parameters $(\widetilde{\alpha_1},\Delta_1)$ and $(\widetilde{\alpha_2},\Delta_2)$, respectively. This results in a maximal feasible transmission rate $\widetilde{R_2}$ given, as before, by $\widetilde{R_2}=\max_{\alpha_2\in[0,1]}I(W_2;r)$. However, the corresponding scale parameter α_2 is set this time to its optimal choice, i.e., $\widetilde{\alpha_2}=\sqrt{P_2/(P_2+2.71(N+P_1))}$.[12] The resulting transmission rate pair $(\widetilde{R_1},\widetilde{R_2})$ is represented by the corner point (B') in Fig. 9 for two examples of channel conditions: weak noise [shown in Fig. 9(b)] and strong noise modeling a strong channel attack on the composite watermark $\mathbf{X}=\mathbf{X}_1+\mathbf{X}_2$ [shown in Fig. 9(a)]. Reversing the roles of the watermarks $\mathbf{X}_1$ and $\mathbf{X}_2$, the joint design also pushes out the corner point (C1') to (C'). More generally, any rate pair on the region frontier delimited by the corner points (A'), (B'), (C'), and (D') is made practically feasible by subsequent time-sharing. When the message W_i *travels alone* over the watermark channel, the equivalent channel is $\mathbf{Y}_i=\mathbf{Y}-\mathbf{U}_j,(i,j)\in\{1,2\}\times\{1,2\},i\neq j$. Hence, W_i can be sent at its maximum feasible rate, which is given by $\max_{\alpha_i\in[0,1]}I(W_i;r_i)$, with $\mathbf{r}_i=\mathcal{Q}_{\Delta_i}(\mathbf{y}_i)-\mathbf{y}_i$. When the two messages *travel together*, the maximal sum of the two feasible rates corresponds to one of the two (say W_1) set to its maximal feasible rate and the other (W_2) facing a total channel noise of $\mathbf{z}+\mathbf{x}_1$. Of course, we can reverse the roles of W_1 and W_2,

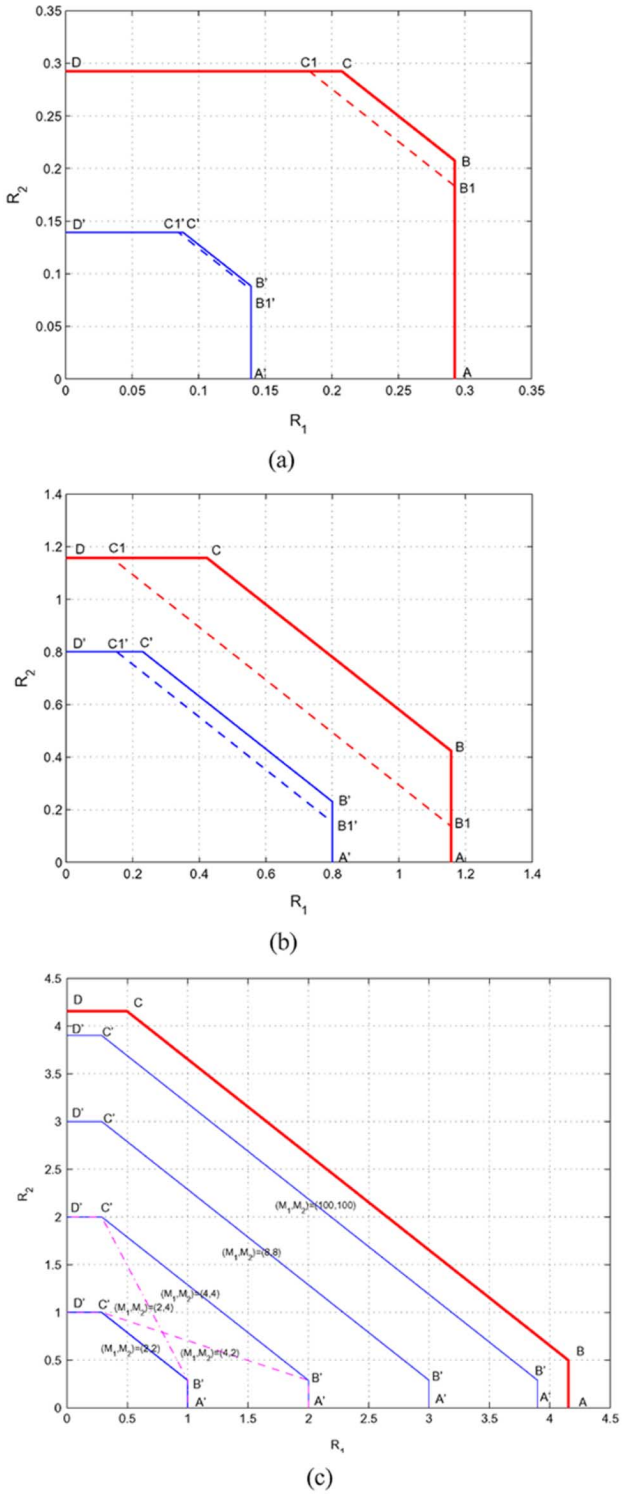


Fig. 9. MAC-aware multiple user information embedding. The improvement brought by "awareness" is depicted for (a) strong channel noise $P_1=P_2$, $(P_1+P_2)/N=0$ dB and (b) weak channel noise $P_1=P_2$, $(P_1+P_2)/N=9$ dB. Solid line delineates the capacity region of the MAC-aware scheme achievable theoretically (upper) and practically (lower). Dashed line delineates the rate region of the MAC-unaware scheme achievable theoretically (upper) and practically (lower). (c) Capacity region of the MAC-aware scheme with (M_1-ary, M_2-ary) input alphabets for very high SNR.

[11]Note, however, that as α_1 depends on α_2, the optimal inflation parameter for DPC1 becomes $\alpha_1=P_1/(P_1+P_2+N)$.

[12]Note that the optimal inflation parameter for SCS1 is $\widetilde{\alpha_1}=(P_1+N)\sqrt{P_1/P_1+2.71N}/(P_1+P_2+N)$.

and the maximal feasible sum rate remains unchanged. Consequently, the achievable rate region $\widetilde{\mathcal{R}}_{\mathrm{MAC}}$ is given by

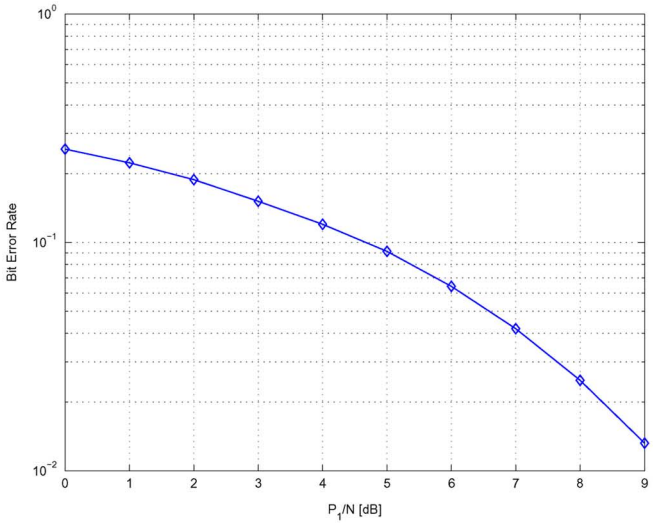
$$\begin{aligned}\widetilde{\mathcal{R}}_{\mathrm{MAC}}(P_1,P_2) &= \Bigg\{(\widetilde{R_1},\widetilde{R_2}):\\ &\widetilde{R_1} \le \max_{\alpha_1\in[0,1]} I\left(W_1;\mathcal{Q}_{\Delta_1(\alpha_1,P_1)}(\mathbf{y}_1)-\mathbf{y}_1\right),\\ &\widetilde{R_2} \le \max_{\alpha_2\in[0,1]} I\left(W_2;\mathcal{Q}_{\Delta_2(\alpha_2,P_2)}(\mathbf{y}_2)-\mathbf{y}_2\right),\\ &\widetilde{R_1}+\widetilde{R_2} \le \max_{\alpha_1\in[0,1]} I\left(W_1;\mathcal{Q}_{\Delta_1(\alpha_1,P_1)}(\mathbf{y}_1)-\mathbf{y}_1\right)\\ &\quad+\max_{\alpha_2\in[0,1]} I\left(W_2;\mathcal{Q}_{\Delta_2(\alpha_2,P_2)}(\mathbf{y})-\mathbf{y}\right)\Bigg\}. \end{aligned} \tag{20}$$

Fig. 9 shows the achievable rate region $\widetilde{\mathcal{R}}_{\mathrm{MAC}}$ gain brought by the joint design of the DPCs in approaching the theoretical limit $\mathcal{C}_{\mathrm{MAC}}$ (18). This improvement, which is more visible at large SNR (i.e., weak channel noise), is more significant in the situations where W_1 and W_2 are both transmitted with nonzero rates. In this case, for a given transmission rate $\widetilde{R_2}$ of W_2, the maximal transmission rate at which W_1 can be sent is larger, and equivalently, for any rate $\widetilde{R_1}$. Moreover the gap to the theoretical limit $\mathcal{C}_{\mathrm{MAC}}$ can be reduced by use of sufficiently large size alphabets $\mathcal{M}_1$ and $\mathcal{M}_2$, as shown in Fig. 9(c). Of course, this is achieved at the cost of a slight increase in encoding and decoding complexities.

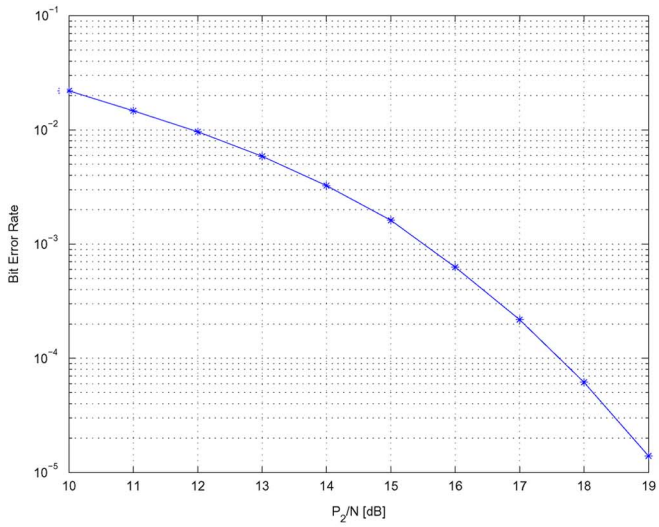
4) BER Analysis and Discussion: Consider the coding scheme given by (14) and (19). The *peeling-off* technique aims to clean up the channel before decoding W_1, by subtracting the codeword $\mathbf{U}_2$. This is good for performance evaluation and for theoretically proving the achievability of the corner point (B) of the capacity region. However, in practice, the decoder does not know the exact codeword $\mathbf{U}_2$ that had been selected at "encoder 2," basically because the host-signal $\mathbf{S}$ is unknown at the receiver. Instead, the decoder determines an estimate $\widehat{\mathbf{U}}_2$ of $\mathbf{U}_2$, as the reconstruction vector of a scaled version of the received signal $\mathbf{Y}$. Of course, the accuracy of this estimation (and, thereby, that of decoding message W_1) depends on the value of SNR2. For instance, bad SNR2 likely causes decoding of W_2 to fail. Thus, the estimate $\widehat{\mathbf{U}}_2$ does not resemble the exact $\mathbf{U}_2$ and it is rather seen as an additional noise source. However, at good (high) SNR2, the estimate $\widehat{\mathbf{U}}_2$ of codeword $\mathbf{U}_2$ is accurate and the *peeling-off* technique is efficient, as shown in Fig. 10. The curves in Fig. 10 are obtained using scalar codebooks and with power allocation such that $P_2 = 10P_1$. Message W_2 is decoded first, corresponding codeword $\hat{U}_2$ is subtracted, and then, message W_1 is decoded. Observe that decoding of message W_1 is more accurate than that of message W_2: For example, observe that $\mathrm{BER}(W_2) \approx 2\times10^{-2}$ at $P_2/N = 10$ dB whereas $\mathrm{BER}(W_1) \approx 1.4\times10^{-2}$ at only $P_1/N = 9$ dB.

V. Multiuser Information Embedding and Structured Lattice-Based Codebooks

In this section, we extend the results obtained in Section IV in the context of two watermarks to the general multiple watermarking case. We also broaden our view to consider the high-dimensional lattice-based codebooks case.



(a)



(b)

Fig. 10. MAC-aware multiple user information embedding BERs using scalar codebooks. The two messages W_1 and W_2 are sent at rates $(\widetilde{R_1},\widetilde{R_2})$ corresponding to the corner point (B') in the capacity region diagram shown in Fig. 9. Successive decoding is implemented by first decoding message W_2, subtracting its corresponding codeword $\hat{\mathbf{u}}_2$, and then, decoding message W_1. (a) Decoding of W_1. (b) Decoding of W_2.

A. Broadcast-Aware Information Embedding: The L-Watermarks Case

The results in Section IV-A can be straightforwardly extended to the situation where, instead of just two messages, L messages $W_i, i = 1,2,\ldots,L$, have to be embedded into the same cover signal $\mathbf{S}$. The composite watermark is $\mathbf{X} = \sum_{i=1}^{L}\mathbf{X}_i$. The watermark $\mathbf{X}_i$ has power P_i and carries the message W_i, where $\sum_{i=1}^{L}P_i = P$. We consider a GBC $\mathbf{Z}_i \sim \mathcal{N}(0,N_i)$ and assume without loss of generality that $N_1 \le N_2 \le \cdots \le N_L$. This means that the watermarks should be designed in such a way that $\mathbf{X}_i$ is less robust than $\mathbf{X}_j$ for $i \le j$. Following the joint DPC scheme, the watermarks should be ordered according to their relative strengths and put on top of each other. This means that the most robust (that is $\mathbf{X}_L$) should be embedded first whereas the most fragile (that is $\mathbf{X}_1$) should be embedded last. For i ranging from L to 1, the watermark signal $\mathbf{X}_i$ is obtained by

applying an $(L-i+1)$th DPC (denoted here by DPCi). The available state information to be used is $\mathbf{S}_i = \mathbf{S} + \sum_{j=i+1}^{L} \mathbf{X}_j$, the sum of the cover signal $\mathbf{S}$ and the already embedded watermarks $\mathbf{X}_j, j > i$. The channel noise is $\mathbf{Z}_i + \sum_{j=1}^{i-1} \mathbf{X}_j$, the sum of the ambient noise $\mathbf{Z}_i$ and the not-yet embedded watermarks $\mathbf{X}_j, j < i$, accumulated and taken as an additional noise component. Note that the Gaussianity of this noise term and its statistic independence from both $\mathbf{X}_i$ and $\mathbf{S}_i$ as well as the statistic independence of $\mathbf{X}_i$ on $\mathbf{S}_i$ conform to the statistical independence between the state information, the watermark and the noise in the original Costa setup [6]. Thus, the optimal inflation parameter for DPCi is $\alpha_i = P_i/(N_i + \sum_{j=1}^{i} P_j)$ and the corresponding maximal achievable rate R_i is given by

$$R_i = \frac{1}{2}\log_2\left(1 + \frac{P_i}{N_i + \sum_{j=1}^{i-1} P_j}\right). \tag{21}$$

A scalar implementation of this broadcast-based joint DPC for embedding L watermarks, consists in L SCSs jointly designed. Similarly to the two-watermark case and using the equivalent channel $\mathbf{y}_i' = \mathbf{y}_i - \sum_{j=i+1}^{L} \mathbf{u}_j$ for SCSi, $i = 1, 2, \ldots, L$, the corresponding achievable rate region is given by the union of all rate L-tuples $(\widetilde{R_1}, \ldots, \widetilde{R_L})$ simultaneously satisfying

$$\widetilde{R_i} \leq \max_{\alpha_i \in [0,1]} I\left(W_i; \mathcal{Q}_{\Delta_i(\alpha_i, P_i)}(\mathbf{y}_i') - \mathbf{y}_i'\right). \tag{22}$$

The union is taken over all power assignments $\{P_i\}, i = 1, 2, \ldots, L$, satisfying the average power constraint $\sum_{j=1}^{L} P_i = P$. The inflation parameter maximizing the right-hand side term of (22) is

$$\widetilde{\alpha_i} = \sqrt{\frac{P_i}{P_i + 2.71\left(N_i + \sum_{j=1}^{i-1} P_j\right)}}. \tag{23}$$

B. MAC-Aware Information Embedding: The K-Watermarks Case

The results in Section IV-B can be straightforwardly extended to the situation where, instead of just two messages, K messages $W_i, i = 1, \ldots, K$, have to be independently encoded into the same cover signal $\mathbf{S}$ and jointly decoded, by the same watermarking authority. We suppose that the watermark $\mathbf{X}_i$, carrying $W_i, i = 1, \ldots, K$, has power P_i. Also, we denote by $\mathbf{Z} \sim \mathcal{N}(0, N)$ the channel noise, assumed to be i.i.d. Gaussian. Functionally, this is a K-user GMAC with state information available at the transmitters but not to the receiver, as argued in Section III-B. The capacity region of such a channel follows a straightforward generalization of (18). This region is given by the union of all rate K-tuples simultaneously satisfying

$$R_i \leq \frac{1}{2}\log_2\left(1 + \frac{P_i}{N}\right), \qquad i = 1, 2, \ldots, K \tag{24a}$$

$$\sum_{j=1}^{K} R_j \leq \frac{1}{2}\log_2\left(1 + \sum_{i=1}^{K} P_i/N\right) \tag{24b}$$

where the union is taken over all power assignments $\{P_i\}, i = 1, \ldots, K$. Following the two-message case considered previously, any corner point of this region can be attained by applying K well-designed DPCs. Consider, for example, the corner point (B) corresponding to the message W_1 transmitted at its maximum rate. Upon reception of $\mathbf{Y} = \sum_{i=1}^{K} \mathbf{X}_i + \mathbf{S} + \mathbf{Z}$, the receiver should perform successive decoding so as to reliably decode the K-tuple $(W_1, W_2, \ldots, W_K)$. In order to attain the corner point (B), decoding should be performed in such a way that W_K is decoded first, W_1 is decoded last, and W_j is decoded before W_i for $j > i$. Consequently, coding consists in a set of K DPCs, denoted by DPCi, with i ranging from K to 1. At the receiver, the decoder sees the equivalent channel $\mathbf{Y} - \sum_{j>i} \mathbf{U}_j$ in the decoding of the message W_i. Thus, an optimal DPCi for this equivalent channel is given by $\mathbf{X}_i = \mathbf{U}_i - \alpha_i \mathbf{S}$ where $\mathbf{U}_i \,|\, \mathbf{S} \sim \mathcal{N}(\alpha_i \mathbf{S}, P_i)$ and $\alpha_i = P_i/(\sum_{j=1}^{K} P_j + N)$. With this theoretical setup, it is possible to reliably transmit all the messages together, with W_i sent at rate $R_i = (1/2)\log_2(1 + P_i/(\sum_{j=1}^{i-1} P_j + N))$. This rate is the maximal rate at which W_i can be transmitted as long as the other messages $W_j, j \neq i$ are simultaneously transmitted at nonzero rates. A scalar implementation of this (K users) GMAC-based joint DPC scheme consists in successively applying K well-designed SCSs. Equivalent channel for SCSi is $\mathbf{y}_{i,b} = \mathbf{y} - \sum_{j=i+1}^{K} \mathbf{u}_j$, which is the received signal assuming interference from only the (i-1) *beforehand* watermarks $\mathbf{x}_j, j < i$ and no *posthand* interference from the remaining $(K-i)$ watermarks $\mathbf{x}_j, j > i$. We also denote by $\mathbf{y}_i \triangleq \mathbf{y}_{i,0} = \mathbf{x}_i + \mathbf{s} + \mathbf{z}$ the received signal assuming neither beforehand nor posthand interferences. The set of feasible rates achieved by this practical coding can be obtained as a straightforward generalization of (20). The corresponding achievable rate region is given by the convex hull of all rate K-tuples $(\widetilde{R_1}, \ldots, \widetilde{R_K})$ simultaneously satisfying

$$\widetilde{R_i} \leq \max_{\alpha_1 \in [0,1]} I\left(W_i; \mathcal{Q}_{\Delta_i}(\mathbf{y}_i) - \mathbf{y}_i\right), \qquad i = 1, 2, \ldots, K \tag{25a}$$

$$\sum_{j=1}^{K} \widetilde{R_j} \leq \sum_{j=1}^{K} \max_{\alpha_j \in [0,1]} I\left(W_j; \mathcal{Q}_{\Delta_j}(\mathbf{y}_{j,b}) - \mathbf{y}_{j,b}\right). \tag{25b}$$

The maximum of the mutual information $I(W_i; \mathcal{Q}_{\Delta_i}(\mathbf{y}_i) - \mathbf{y}_i)$ is attained with the optimal choice of $\alpha_i \in [0, 1]$ given by

$$\widetilde{\alpha_i} = \left(1 - \sum_{j=i+1}^{K} \alpha_j\right)\sqrt{\frac{P_i}{P_i + 2.71N}}$$

with

$$\widetilde{\alpha_K} = \sqrt{\frac{P_K}{P_K + 2.71N}}.$$

C. Lattice-Based Codebooks for BC-Aware Multiuser Information Embedding

The gap to the ideal capacity region of the sample-wise joint scalar DPC practical capacity region shown in Fig. 6 can be partially bridged using structured finite-dimensional lattice-based codebooks. Lattices have been proposed in the context of multiterminal binning in [24] and have been considered for the first time in the context of single-user watermarking in [13]. Consequent works [14]–[16] extended these results to different scenarios. In what follows, only the required ingredients are briefly reviewed. The reader may refer to [25] for a full discussion.

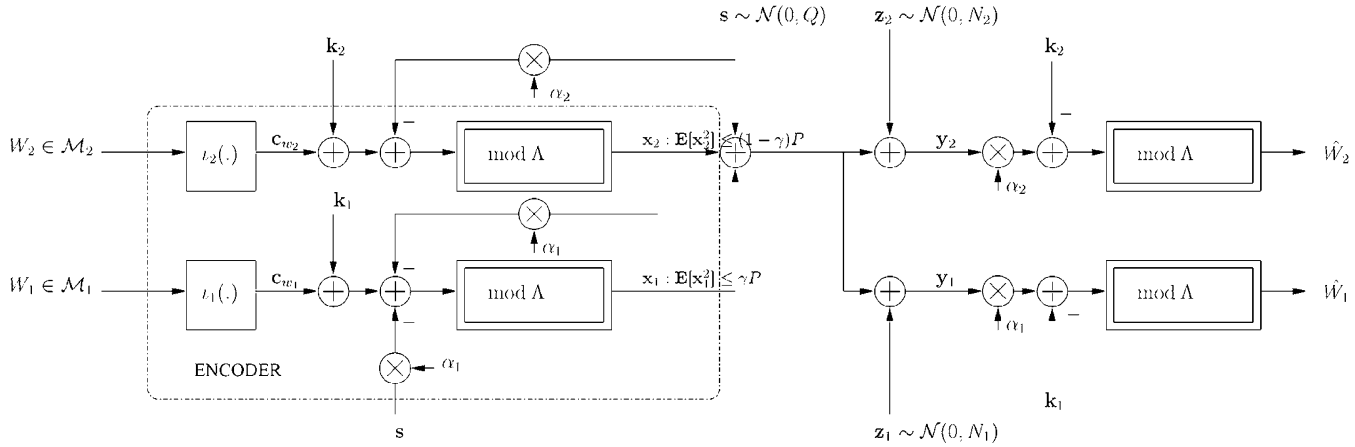


Fig. 11. Lattice-based scheme for multiple information embedding over a GBC.

Consider the transmission scheme depicted in Fig. 11 where Λ is some n-dimensional lattice. This scheme is a generalization to the lattice codebook case of a slight variation of the scalar case considered in Section IV-A.[13] The function $\ell_1(\cdot)$ is used for arbitrary mapping the set of indexes $W_1 \in \mathcal{M}_1 = \{1, \ldots, M_1\}$ to a certain set of vectors $\mathcal{C}_{w_1} = \{\mathbf{c}_{w_1} \quad w_1 = 1, \ldots, M_1\}$ to be specified in the sequel. The function $\ell_2(.)$ does similarly for the set of indexes $W_2 \in \mathcal{M}_2 = \{1, \ldots, M_2\}$. With respect to the scalar codebook case, $\mathcal{C}_{w_i}, i = 1, 2$, is a lattice codebook whose entries must be appropriately chosen so as to maximize the encoding performance. For each $W_i \in \mathcal{M}_i$, with $i = 1, 2$, the codeword $\ell_i(W_i) = \mathbf{c}_{w_i}$ is the *coset leader* of the coset $\Lambda_{w_i} = \mathbf{c}_{w_i} + \Lambda$ relative to the lattice Λ. The codebook $\mathcal{C}_{w_i}$ is shared between the encoder and the decoder i and is assumed to be uniformly distributed over the fundamental cell $\mathcal{V}(\Lambda)$ of the lattice Λ. Also, we assume *common randomness*, meaning that the key $\mathbf{k}_i, i = 1, 2$, is known to both the encoder and the decoder i. Apart from obvious security purposes, these keys will turn out to be useful in attaining the capacity region.

In the following, we consider cover signal vectors (frames) of length n. Following (3), the encoding and decoding functions for the lattice-based joint DPC given by (5) and (10) write

$$\mathbf{x}_2(\mathbf{s}; W_2, \Lambda) = (\mathbf{c}_{w_2} + \mathbf{k}_2 - \alpha_2 \mathbf{s}) \mod \Lambda \tag{26a}$$

$$\mathbf{x}_1(\mathbf{s}; W_1, \Lambda) = (\mathbf{c}_{w_1} + \mathbf{k}_1 - \alpha_1(\mathbf{s} + \mathbf{x}_2)) \mod \Lambda \tag{26b}$$

$$\widehat{W_i} = \operatorname{argmin}_{W_i \in \mathcal{M}_i} \times \|(\alpha_i \mathbf{y}_i - \mathbf{k}_i - \mathbf{c}_{w_i}) \bmod \Lambda\|, \qquad i = 1, 2. \tag{26c}$$

The modulo reduction operation is defined as $\mathbf{x} \bmod \Lambda \triangleq \mathbf{x} - \mathcal{Q}_\Lambda(\mathbf{x}) \in \mathcal{V}(\Lambda)$ where the n-dimensional quantization operator $\mathcal{Q}_\Lambda(.)$ is such that quantization of $\mathbf{x} \in \mathbb{R}^n$ results in the closest lattice point $\boldsymbol{\lambda} \in \Lambda$ to $\mathbf{x}$.

We focus on the practically feasible rate region achieved by (26). To this end, we rely on a previous work relative to practical achievable rates with lattice codebooks in the context of a single-user watermark [16], [15]. Here, the situation is different since two watermarks are concerned, but the key ideas remain the same. Thus, details are skipped and we only mention the key steps, in processing the received signals $\mathbf{y}_1$ and $\mathbf{y}_2$. Each of the channels $\mathbf{Y}_1$ and $\mathbf{Y}_2$ is similar to the one in [16] and [15], with, however, a different state information and channel noise. The establishment of the following results relies principally on the properties of a modulo lattice additive noise (MLAN) channel [26] and on the following two important properties of the mod-Λ operation:

P1)

$$\forall(\boldsymbol{\lambda}, \mathbf{a}) \in \Lambda \times \mathbb{R}^n, (\mathbf{a} + \mathbf{v} + \boldsymbol{\lambda}) \bmod \Lambda = (\mathbf{a} + \mathbf{v}) \bmod \Lambda; \tag{27a}$$

P2)

$$\forall(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{2n}, ((\mathbf{x} \bmod \Lambda) + \mathbf{y}) \bmod \Lambda = (\mathbf{x} + \mathbf{y}) \bmod \Lambda. \tag{27b}$$

Upon reception of $\mathbf{y}_i, i = 1, 2$, "receiver i" computes the signal $\mathbf{r}_i = (\alpha_i \mathbf{y}_i - \mathbf{k}_i) \bmod \Lambda$. Using P1) and P2) and straightforward algebra calculations, it can be shown that

$$\mathbf{r}_1 = (\mathbf{c}_{w_1} + \alpha_1 \mathbf{z}_1 - (1 - \alpha_1)\mathbf{x}_1) \bmod \Lambda \tag{28a}$$

$$\mathbf{r}_2 = (\mathbf{c}_{w_2} + \alpha_2(\mathbf{z}_2 + \mathbf{x}_1) - (1 - \alpha_2)\mathbf{x}_2) \bmod \Lambda. \tag{28b}$$

Hence, the "degraded user" (more noisy watermarked content) sees the equivalent channel noise $\widetilde{\mathbf{V}_2} = (\alpha_2(\mathbf{Z}_2 + \mathbf{X}_1) - (1 - \alpha_2)\mathbf{X}_2) \bmod \Lambda$ and the "better user" (less noisy watermarked content) sees the equivalent channel noise $\widetilde{\mathbf{V}_1} = (\alpha_1 \mathbf{Z}_1 - (1 - \alpha_1)\mathbf{X}_1) \bmod \Lambda$. Now, using the important *inflated lattice lemma* reported in [27], $\mathbf{Y}_1$ and $\mathbf{Y}_2$ turn to be two MLAN channels with channel noises $\widetilde{\mathbf{V}_1}$ and $\widetilde{\mathbf{V}_2}$, respectively. The MLAN channel has been first considered in [28] and [29]. It is shown that when modulo reduction is with respect to some lattice Λ and when the channel noise $\mathbf{V}$ is i.i.d. Gaussian, capacity in bits per dimension can be written as

$$C(\Lambda) = \frac{1}{n}\left(\log_2(V(\Lambda)) - h(\mathbf{V})\right) \tag{29}$$

where $h(\cdot)$ denotes differential entropy. Hence, the practically achievable rates $R_1(\Lambda)$ and $R_2(\Lambda)$ are given by (29), with the channel noise $\mathbf{V}$ being replaced by $\widetilde{\mathbf{V}_1}$ and $\widetilde{\mathbf{V}_2}$, respectively. The maximally achievable rates are obtained by maximizing these expressions over α_1 and α_2, respectively. The corre-

[13] More precisely, this is a generalization to the lattice case of a DC-QIM-based two-users watermarking scheme. DC-QIM is considered because it is more convenient and also it has very close performance to SCS as has been reported in Section II-B.

TABLE I
LATTICES WITH THEIR IMPORTANT PARAMETERS

Lattice	Name	n	$G(\Lambda)$	$\gamma_s(\Lambda)$ [dB]	$\gamma_s(\Lambda)$ [bit per dimension]
$\mathbb{Z}$	Integer Lattice	1	$\frac{1}{12}$	0.00	0.000
A_2	Hexagonal Lattice	2	$\frac{5}{36\sqrt{3}}$	0.17	0.028
D_4	$4D$ Checkerboard L.	4	0.0766	0.37	0.061

sponding achievable rate region $\bar{\mathcal{R}}_{\rm BC}$ is given by

$$\bar{\mathcal{R}}_{\rm BC}(P) = \bigcup_{0\le\gamma\le 1}\Bigg\{(\widetilde{R}_1,\widetilde{R}_2):\ \widetilde{R}_1 \le \max_{\alpha_1\in[0,1]}\frac{1}{n}\Bigg(\log_2(V(\Lambda)) - h\Big(\widetilde{\mathbf{V}}_1(\alpha_1,\gamma)\Big)\Bigg),\ \widetilde{R}_2 \le \max_{\alpha_2\in[0,1]}\frac{1}{n}\Bigg(\log_2(V(\Lambda)) - h\Big(\widetilde{\mathbf{V}}_2(\alpha_2,\gamma)\Big)\Bigg)\Bigg\}. \quad (30)$$

Note that from the right-hand side term of (30), we have $\bar{\mathcal{R}}_{\rm BC}\subset\mathcal{C}_{\rm BC}$, where $\mathcal{C}_{\rm BC}$ is the full capacity region of a Gaussian BC with state information at the encoder (9). In general, no closed form of (30) can be derived and the optimal pair (α_1,α_2) has to be computed numerically to evaluate the differential entropy $h(\widetilde{\mathbf{V}}_i), i=1,2$. However, closed form approximations can be found in some special situations as shown hereafter.

1) As the dimensionality n of the lattice goes to infinity, the pdfs of the noises $\widetilde{\mathbf{V}}_1$ and $\widetilde{\mathbf{V}}_2$ tend to Gaussian distributions as quantization errors with respect to this lattice. Consequently, the optimal inflation parameters α_1 and α_2 minimizing $h(\widetilde{\mathbf{V}}_1)$ and $h(\widetilde{\mathbf{V}}_2)$ are those which minimize the variances of $\widetilde{\mathbf{V}}_1$ and $\widetilde{\mathbf{V}}_2$, respectively. These are $\alpha_1 = \gamma P/(\gamma P+N_1)$ and $\alpha_2=(1-\gamma)P/(P+N_2)$. The ideal capacity region is attained with such a choice.
2) For finite-dimension lattice reduction, however, the pdfs of $\widetilde{\mathbf{V}}_1$ and $\widetilde{\mathbf{V}}_2$ are not strictly Gaussian, but rather the convolution of a Gaussian with a uniform distribution. The equality $(\alpha_1,\alpha_2) = ((\gamma P/\gamma P+N_1),((1-\gamma)P/N_2+P))$ does not hold strictly but remains a quite accurate approximation. Considering this approximation leads to $\mathbb{E}_{\widetilde{\mathbf{V}}_1}[\widetilde{\mathbf{V}}_1^2]=\alpha_1 N_1$ and $\mathbb{E}_{\widetilde{\mathbf{V}}_2}[\widetilde{\mathbf{V}}_2^2]=\alpha_2(N_2+\gamma P)$. Now, given that[14] $h(\widetilde{\mathbf{V}}_1)\le \log(2\pi e\alpha_1 N_1)$ and $h(\widetilde{\mathbf{V}}_2)\le\log 2\pi e\alpha_2(N_2+\gamma P)$, we get

$$R_1(\Lambda)\ge\frac{1}{n}\left(\frac{1}{2}\log\left(1+\frac{\gamma P}{N_1}\right)-\frac{1}{2}\log 2\pi e G(\Lambda)\right) \quad (31a)$$

$$R_2(\Lambda)\ge\frac{1}{n}\left(\frac{1}{2}\log\left(1+\frac{(1-\gamma)P}{N_2+\gamma P}\right)-\frac{1}{2}\log 2\pi e G(\Lambda)\right). \quad (31b)$$

This means that by using appropriate lattices for modulo-reduction, we are able to make the gap to the full theoretical capacity region smaller than $\log 2\pi e G(\Lambda)$. This can be achieved by selecting lattices that have good quantization properties. These are those for which the normalized second moment $G(\Lambda)$ approaches $1/2\pi e$.

[14]This is because the normal distribution is the one that maximizes entropy for a given second moment.

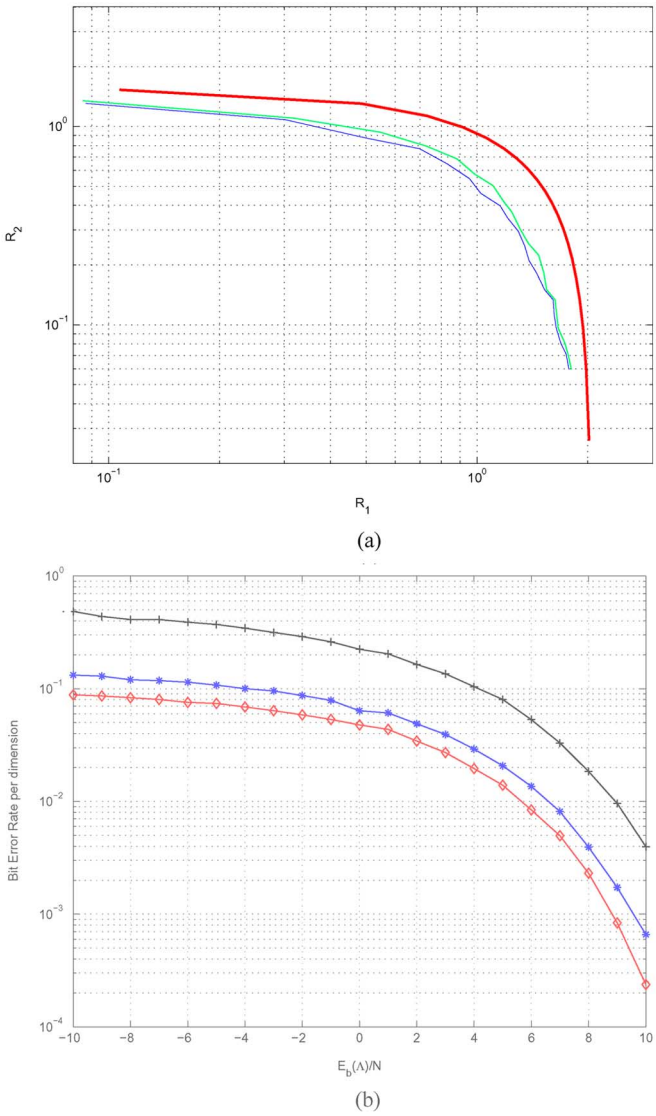


Fig. 12. Performance improvement in multiple user information embedding rates and BER due to the use of lattice codebooks. Coset leaders $\{\mathbf{c}_{w_1}\}$ are chosen among lattice deep holes. (a) Achievable rate region for BC-like multiple user information embedding and (b) corresponding BERs corresponding to the transmission of message W_1. From bottom to top: lattices checkerboard D_4, hexagonal A_2, and cubic $\mathbb{Z}$. (a) Achievable rate region with lattices $\mathbb{Z}$ and A_2. (b) BERs with lattices $\mathbb{Z}$ and A_2 and D_4.

The n-dimensional lattices considered for Monte Carlo achievable rate region integration are summarized in Table I, together with their most important parameters. Achievable rate region curves in bits per dimension are plotted in Fig. 12(a) where we observe that the use of the hexagonal lattice A_2, for example,

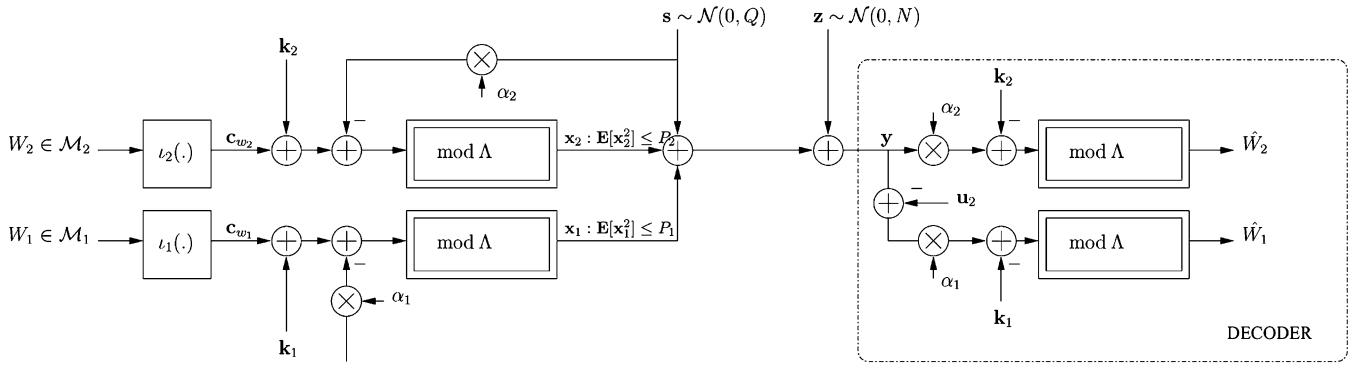


Fig. 13. Lattice-based scheme for multiple information embedding over a GMAC.

enlarges the set of the rate pairs practically feasible, with respect to the scalar lattice $\mathbb{Z}$. Of course, this improvement goes along with a slight increase in computational cost. The same improvement can be observed through BER enhancement visible in Fig. 12(b). Note that Fig. 12(b) only shows the BER corresponding to the transmission of message W_1 (the codebook $\mathcal{C}_{w_1}$ is set to a subset[15] of lattice deep holes [25]). Also, note that for fair comparison of the error correction capability of the different lattices, BER are plotted against the per-bit per-dimension SNR (i.e., energy $E_b(\Lambda)$ needed to transmit one bit per-dimension to noise ratio). The BER curves corresponding to the transmission of message W_2 can be obtained by shifting to the right those of W_1 by the factor $\beta_{\mathrm{BC}}(R_1, R_2) = (R_1/R_2)\times((1-\gamma)P/\gamma P)\times(N_1/\gamma P + N_2)$ (decibels).

D. Lattice-Based Codebooks for MAC-Aware Multiuser Information Embedding

The gap to the capacity region $\mathcal{C}_{\mathrm{MAC}}$ (18) of the achievable rate region $\widetilde{R}_{\mathrm{MAC}}$ (20) shown in Fig. 9 and corresponding to the sample-wise joint scalar DPC can be partially bridged using finite-dimensional lattice-based codebooks. The resulting transmission scheme is depicted in Fig. 13 where Λ is some n-dimensional lattice. The functions $\ell_i(\cdot), i = 1, 2$ and the lattice codebooks $\mathcal{C}_{w_i}, i = 1, 2$ are defined in a similar way to that in the broadcast case addressed previously. We focus on the improvement of the feasible rate pair $(R_1(\Lambda), R_2(\Lambda))$ brought by the use of the lattice codebooks $\mathcal{C}_{w_i}, i = 1, 2$, with comparison to the baseline scalar codebooks considered in Section IV-B. Consider, for example, the corner point (B') of the capacity region shown in Fig. 9. The encoding and decoding of W_1 and W_2 are performed according to

$$\begin{aligned}
\mathbf{x}_1(\mathbf{s}; W_1, \Lambda) &= (\mathbf{c}_{w_1} + \mathbf{k}_1 - \alpha_1(1-\alpha_2)\mathbf{s}) \bmod \Lambda \\
\mathbf{x}_2(\mathbf{s}; W_2, \Lambda) &= (\mathbf{c}_{w_2} + \mathbf{k}_2 - \alpha_2\mathbf{s}) \bmod \Lambda \\
\widehat{W_1} &= \operatorname{argmin}_{W_1 \in \mathcal{M}_1} \|(\alpha_1\mathbf{y}_1 - \mathbf{k}_1 - \mathbf{c}_{w_1}) \bmod \Lambda\| \\
\widehat{W_2} &= \operatorname{argmin}_{W_2 \in \mathcal{M}_2} \|(\alpha_2\mathbf{y} - \mathbf{k}_2 - \mathbf{c}_{w_2}) \bmod \Lambda\|
\end{aligned} \tag{32}$$

where $\mathbf{y}_1 = \mathbf{y} - (\mathbf{x}_2 + \alpha_2\mathbf{s})$. Upon reception, the receiver first computes the error signal $\mathbf{r} = (\alpha\mathbf{y} - \mathbf{k}_2) \bmod \Lambda$. In a similar way to that for the broadcast case, it can be shown that $\mathbf{r} = (\mathbf{c}_{w_2} + \alpha_2(\mathbf{z} + \mathbf{x}_1) - (1-\alpha_2)\mathbf{x}_2) \bmod \Lambda$. Hence, the equivalent channel for the transmission of W_2 is an MLAN channel with (Gaussian) channel noise $\widetilde{\mathbf{v}_2} = (\alpha_2(\mathbf{z} + \mathbf{x}_1) - (1-\alpha_2)\mathbf{x}_2) \bmod \Lambda$. Next, the receiver computes $\mathbf{r}_1 = (\alpha\mathbf{y}_1 - \mathbf{k}_1) \bmod \Lambda$, which can be shown to equal $(\mathbf{c}_{w_1} + \alpha_1\mathbf{z} - (1-\alpha_1)\mathbf{x}_1) \bmod \Lambda$, completely independent of $\mathbf{x}_2$. Hence, the equivalent channel for the transmission of W_1 is another MLAN channel with (Gaussian) channel noise $\widetilde{\mathbf{v}_1} = (\alpha_1\mathbf{z} - (1-\alpha_1)\mathbf{x}_1) \bmod \Lambda$. Consequently, by using (32), the achievable rate pair $(R_1(B'), R_2(B'))$ corresponding to the corner point (B') of the capacity region $\mathcal{C}_{\mathrm{MAC}}$ is given by

$$R_1(B') = \max_{\alpha_1 \in [0,1]} \frac{1}{n}\left(\log_2(V(\Lambda)) - h\left(\widetilde{\mathbf{V}}_1(\alpha_1, P_1)\right)\right) \tag{33a}$$

$$R_2(B') = \max_{\alpha_2 \in [0,1]} \frac{1}{n}\left(\log_2(V(\Lambda)) - h\left(\widetilde{\mathbf{V}}_2(\alpha_2, P_2)\right)\right). \tag{33b}$$

Note that $(R_1, R_2) \in \mathcal{C}_{\mathrm{MAC}}$. Similarly to the development made in the broadcast case, the achievable rate region by using the modulo reduction with respect to the lattice Λ straightforwardly generalizes (20) and it is given by

$$\begin{aligned}
&\bar{\mathcal{R}}_{\mathrm{MAC}}(P_1, P_2) \\
&\quad = \Bigg\{(\widetilde{R}_1, \widetilde{R}_2): \\
&\qquad \widetilde{R}_1 \leq \max_{\alpha_1 \in [0,1]} \frac{1}{n}\left(\log_2(V(\Lambda)) - h\left(\widetilde{\mathbf{V}}_1(\alpha_1, P_1)\right)\right) \\
&\qquad \widetilde{R}_2 \leq \max_{\alpha_2 \in [0,1]} \frac{1}{n}\left(\log_2(V(\Lambda)) - h\left(\widetilde{\mathbf{V}}_2(\alpha_2, P_2)\right)\right) \\
&\qquad \widetilde{R}_1 + \widetilde{R}_2 \leq \max_{\alpha_1 \in [0,1]} \frac{1}{n}\left(\log_2(V(\Lambda)) - h\left(\widetilde{\mathbf{V}}_1(\alpha_1, P_1)\right)\right) \\
&\qquad\qquad + \max_{\alpha_2 \in [0,1]} \frac{1}{n}\left(\log_2(V(\Lambda)) - h\left(\widetilde{\mathbf{V}}(\alpha_2, P_2)\right)\right)\Bigg\}
\end{aligned} \tag{34}$$

where $\widetilde{\mathbf{V}_i} = (\alpha_i\mathbf{Z} - (1-\alpha_i)\mathbf{X}_i) \bmod \Lambda, i = 1, 2$ and $\tilde{\mathbf{V}} = (\alpha_2(\mathbf{Z} + \mathbf{X}_1) - (1-\alpha_2)\mathbf{X}_2) \bmod \Lambda$.

The improvement in BER brought by lattice coding in decoding message W_1 is illustrated in Fig. 12(b). As in the broadcast case, the BER curves corresponding to the transmission of message W_2 can be obtained by translating to the right those of W_1, by $\beta_{\mathrm{MAC}}(R_1, R_2) = (R_1P_2N/R_2P_1(N + P_1))$ (decibels).

[15]Note that since the curves in Fig. 12(b) are plotted against the per-bit per-dimension SNR, results do not depend on the number of lattice holes used (i.e., the transmission rate).

VI. Conclusion

In this paper, we first investigated the tight relationship between multiple user information embedding and conventional multiuser information theory. For instance, two different situations of embedding several messages into one common cover signal are emphasized. The first situation is recognized as being equivalent to communication over a GBC with state information noncausally known at the transmitter but not at the receivers. The second is argued as to be analog to communication over a GMAC with state information known noncausally at the transmitters but not at the receiver. Next, based on this equivalence and heavily relying on a recent work by Kim *et al.* [19] in which the authors extend the single-user Costa's DPC to the multiuser case, two practically feasible scalar schemes for simultaneously embedding two messages into the same host signal are proposed. These schemes carefully extend the initial QIM and SCS schemes, that were originally conceived for embedding one watermark, to the two-watermark case. The careful design concerns the joint encoding as well as the appropriate order needed so as to reliably embed the different watermarks. A central idea for the joint design is "awareness." The improvement brought by this awareness is shown through comparison to the corresponding rather intuitive schemes, obtained through superimposition, as many times as needed, of the single-user schemes QIM and SCS. Performance is analyzed in terms of both achievable rate region and BER. Finally, the proposed schemes are straightforwardly extended to the arbitrary number of watermarks case and also to the vector case through lattice-based codebooks. Results are supported by illustrative achievable rate region and BER curves obtained through Monte Carlo integration and Monte Carlo simulation, respectively.

References

[1] A. Zaidi, P. Piantanida, and P. Duhamel, "Scalar scheme for multiple user information embedding," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Philadelphia, PA, Mar. 18–23, 2005, pp. 5–8.

[2] I. Cox, M. Miller, and A. McKellips, "Electronic watermarking: The first 50 years," in *Proc. Int. Workshop Multimedia Signal Process.*, 2001, pp. 225–230.

[3] B. Chen and G. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1423–1443, May 2001.

[4] I. Cox, M. Miller, and A. McKellips, "Watermarking as communication with side information," in *Proc. Int. Conf. Multimedia Comput. Syst.*, Jul. 1999, pp. 1127–1141.

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

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

[7] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems Control IT.*, vol. 9, pp. 19–31, 1980.

[8] C. D. Heegard and A. A. E. Gamal, "On the capacity of computer memory with defects," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 5, pp. 731–739, Sep. 1983.

[9] R. Tzschoppe, R. Bäuml, R. Fischer, J. Huber, and A. Kaup, "Additive non-Gaussian noise attacks on the scalar Costa scheme (SCS)," in *Proc. SPIE 1ST*, San Joze, CA, Jan. 2005, pp. 114–123.

[10] N. Liu and K. P. Subbalakshmi, "Non-uniform quantizer design for image data hiding," in *Proc. IEEE Int. Conf. Image Process.*, Singapore, Oct. 2004, vol. 4, pp. 2179–2182.

[11] S. Voloshynovskiy, O. Koval, F. Prez-Gonzáles, M. K. Mihcak, J. E. Vila-Forcen, and T. Pun, "Data-hiding with partially available side information," in *13th Eur. Signal Process. Conf.*, Antalya, Turkey, Sep. 4–8, 2005 [Online]. Available: http://vision.unige.ch/publications/postscript/2005/VoloshynovskiyKovalPerezGonzalezMihcakPun_SP2005.pdf).

[12] A. Zaidi and P. Duhamel, "On coding with a partial knowledge of the state information," in *Proc. Asilomar Conf. Signals, Systems Comput.*, Monterey, CA, Oct. 2005, pp. 657–661.

[13] R. F. H. Fischer, R. Tzschoppe, and R. Bäeuml, "Lattice costa schemes using subspace projection for digital watermarking," in *Proc. ITG Conf. Source Channel Coding*, Erlangen, Germany, Jan. 2004, pp. 127–134.

[14] P. Moulin and R. Koetter, "Data-hiding codes," *Proc. IEEE*, vol. 93, no. 12, pp. 2083–2126, Dec. 2005.

[15] A. Zaidi and P. Duhamel, "Joint source-channel coding for lattice watermarking," in *Proc. Euro. Signal Process. Conf.*, Antalya, Turkey, Sep. 2005, CD-ROM.

[16] ——, "Modulo lattice additive noise channel for QIM watermarking," in *Proc. Int. Conf. Image Process.*, Geneva, Italy, Sep. 2005, pp. 993–996.

[17] A. Zaidi and P. Duhamel, "On capacity and bit error rate computation in finite-dimensional lattice coding for dirty paper coding," *IEEE Trans. Signal Process.*, Mar. 2007, submitted for publication.

[18] A. Zaidi, P. Piantanida, and P. Duhamel, "Mac-aware coding strategy for multiple user information embedding," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Toulouse, France, Oct. 21, 2005.

[19] Y.-H. Kim, A. Sutivong, and S. Sigurjonsson, "Multiple user writing on dirty paper," in *Proc. ISIT* , Chicago, IL, Jun. 2004, p. 534.

[20] B. Chen and G. Wornell, "Achievable performance of digital watermarking systems," in *Proc. Int. Conf. Multimedia Comput. Syst.*, Florence, Italy, Jun. 1999, vol. 87, pp. 13–18.

[21] T. M. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 2–14, Jan. 1972.

[22] T. M. Cover, "Comments on broadcast channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2524–2530, Oct. 1988.

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

[24] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.

[25] J. H. Conway and N. J. A. Sloane, *Sphere Packing, Lattices and Groups*, 3rd ed. New York: Wiley, 1988.

[26] G. D. Forney, M. D. Trott, and S. Y. Chung, "Sphere-bound-achieving cosets codes and multilevel coset codes," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 820–850, May 2000.

[27] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for cancelling known interference," in *Int. Symps. Inf. Theory Appl.*, Honolulu, HI, 2000, pp. 681–684.

[28] G. D. Forney and L. F. Wei, "Multidimensional constellations—Part I: Introductions figures of merit, and generalized cross constellations," *IEEE J. Select. Areas Commun.*, vol. 7, no. 6, pp. 877–892, Aug. 1989.

[29] J. G. D. Forney, "Multidimensional constellations—Part II: Voronoi constellations," *IEEE J. Select. Areas Commun.*, vol. 7, no. 6, pp. 941–958, Aug. 1989.

Abdellatif Zaidi was born in Tunisia in 1978. He received the Eng. degree in electrical engineering from the High School of Advanced Techniques (ENSTA) Paris, France, in 2002 and the M.Sc. and Ph.D. degrees from National School of Telecommunication (ENST), Paris, France, in 2002 and 2005, respectively.

From December 2002 to December 2005, he was with the Communications an Electronics Department (COM/ELEC), ENST, and the Signals and Systems Laboratory (LSS/CNRS), Gif-Sur-Yvette, France. Currently, he is a Research Assistant at the Communications and Remote Sensing Laboratory (TELE), Université catholique de Louvain (UCL), Louvain-la-Neuve, Belgium. His research interests cover a broad angle of topics from signal processing for communication and multiuser information theory. Of particular emphasis is the problem of coding for side-informed channels. This includes the questions of coding and interference mitigation in multiuser channels, and more recently, relaying problems and cooperative communication, with application to sensor networking and ad hoc wireless networks.

Pablo Piantanida was born in Argentina in 1978. He received the Eng. degree in electrical engineering from the Faculty of Engineering, the University of Buenos Aires, Buenos Aires, Argentina. Currently, he is working towards the Ph.D. degree at the CNRS/LSS (Laboratoire de Signaux et Systemes, Gif sur Yvette, France), from Orsay University, Orsay, France, where he is developing studies in Shannon theory and related problems in multiuser information theory.

From 2000 to 2003, he was with the Institute of Biomedical Engineering at the University of Buenos Aires, where his research interests were in speech recognition and language modeling.

Pierre Duhamel (F'98) was born in France in 1953. He received the Eng. degree in electrical engineering from the National Institute for Applied Sciences (INSA), Rennes, France, in 1975, and the Dr. Eng. and Doctoratès sciences degrees from Orsay University, Orsay, France, in 1978 and 1986, respectively.

From 1975 to 1980, he was with Thomson-CSF, Paris, France, where his research interests were in circuit theory and signal processing, including digital filtering and analog fault diagnosis. In 1980, he joined the National Research Center in Telecommunications (CNET), Issy les Moulineaux, France, where his research activities were first concerned with the design of recursive CCD filters. Later, he worked on fast algorithms for computing Fourier transforms and convolutions, and applied similar techniques to adaptive filtering, spectral analysis, and wavelet transforms. From 1993 to September 2000, he was a Professor at National School of Engineering in Telecommunications (ENST), Paris, France, with research activities focused on signal processing for communications. He was the Head of the Signal and Image Processing Department from 1997 to 2000. He is now with Laboratoire de Signaux et Systemes (CNRS/LSS), Gif-sur-Yvette, France, where he is developing studies in signal processing for communications (including equalization, iterative decoding, multicarrier systems) and signal/image processing for multimedia applications, including source coding, joint source/channel coding, watermarking, and audio processing.

Dr. Duhamel was Chairman of the DSP Committee from 1996 to 1998, and member of the SP for Com committee until 2001. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1989 to 1991, an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS, and a Guest Editor for the special issue of the IEEE TRANSACTIONS ON SIGNAL PROCESSING on wavelets. He was a Distinguished Lecturer in 1999, and Cogeneral Chair of the 2001 International Workshop on Multimedia Signal Processing, Cannes, France. He was also Cotechnical Chair of ICASSP 2006, Toulouse, France. The paper on subspace-based methods for blind equalization, which he coauthored, received the Best Paper Award from the IEEE TRANSACTIONS ON SIGNAL PROCESSING in 1998. He was awarded the "grand prix France Telecom" by the French Science Academy in 2000.