Scalar Costa Scheme for Information Embedding

Joachim J. Eggers, Robert Bäuml, Roman Tzschoppe, Bernd Girod

Abstract

[1] Research on information embedding and particularly information hiding techniques has received considerable attention within the last years due to its potential application in multimedia security. Digital watermarking, which is an information hiding technique where the embedded information is robust against malicious or accidental attacks, might offer new possibilities to enforce the copyrights of multimedia data. In this article, the specific case of information embedding into independent identically distributed (IID) data and attacks by additive white Gaussian noise (AWGN) is considered. The original data is not available to the decoder. For Gaussian data, Costa proposed already in 1983 a scheme that theoretically achieves the capacity of this communication scenario. However, Costa's scheme is not practical. Thus, several research groups have proposed suboptimal practical communication schemes based on Costa's idea. The goal of this artical is to give a complete performance analysis of the *scalar Costa scheme* (SCS) which is a suboptimal technique using scalar embedding and reception functions. Information theoretic bounds and simulation results with state-of-the-art coding techniques are compared. Further, reception after amplitude scaling attacks and the invertibility of SCS embedding are investigated.

Keywords

Information embedding, communication with side-information, blind digital watermarking, scalar Costa scheme

I. Introduction

RESEARCH on information embedding has gained substantial attention during the last years. This is mainly due to the increased interest in digital watermarking technology which potentially can solve copyright infringements and data integrity disputes. Digital watermarking is considered as the *imperceptible, robust, secure communication* of information by embedding it in and retrieving it from other digital data. The basic idea is that the embedded information – the watermark message – travels with the multimedia data wherever the watermarked data goes. Over the last years, many different watermarking schemes for a large variety of data types have been developed. Most of the work considers still image

[1] submitted to: IEEE Transactions on Signal Processing, Special Issue on: Signal Processing for Data Hiding in Digital Media & Secure Content Delivery

data, but watermarking of audio and video data is popular as well. Theoretical limits of digital watermarking have been investigated since about 1999 [1], [2], [3]. In general, watermark embedding techniques and attacks against watermarks have to be designed specifically for certain host data types. A particularly interesting case is that of independent identically distributed (IID) host data and attacks by additive white Gaussian noise (AWGN). The analysis of more complicated scenarios can often be ascribed to this special case [4], [5], [6].

In this paper, we focus on information embedding into IID host data facing AWGN attacks. Throughout the paper, we denote the investigated scenario as a watermarking scenario. However, it should be emphasized that AWGN is not the optimum attack against embedded watermarks for all types of host data. Thus, *information embedding* might be a more correct term than *digital watermarking*, since the robustness requirement is weakenend to robustness against AWGN. We further constrain the discussion to *blind reception*, meaning that the decoder has no access to the original data.

Fig. 1 depicts a block diagram of the considered blind watermarking scenario. A *watermark message* m is embedded into IID *original data* $\mathbf{x}$ of power σ_{x}^2 to produce the *watermarked data* $\mathbf{s}$. The difference $\mathbf{w} = \mathbf{s} - \mathbf{x}$ is denoted the *watermark signal.* Here, we consider only embedding techniques giving a zero-mean watermark signal $\mathbf{w}$ with power limited to σ_{w}^2. Next, AWGN $\mathbf{v}$ of power σ_{v}^2 is added to the watermarked data $\mathbf{s}$. This process, denoted *AWGN attack*, produces the *attacked data* $\mathbf{r}$. The attacked data $\mathbf{r}$ is identical to the *received data* $\mathbf{r}$, which is input to the watermark decoder. The embedding process and decoding process is dependent on the key K to achieve security of the communication. Usually, a key sequence $\mathbf{k}$ with the same length as $\mathbf{x}$ is derived from the key K. In this paper, $\mathbf{x}$,$\mathbf{w}$,$\mathbf{s}$,$\mathbf{r}$, and $\mathbf{k}$ are vectors of identical length L_x, and x_n,w_n,s_n,r_n, and k_n refer to their respective nth elements. Random variables are written in Sans Serif font, e.g., x for a scalar random variable and $\boldsymbol{\mathsf{x}}$ for a vector random variable.

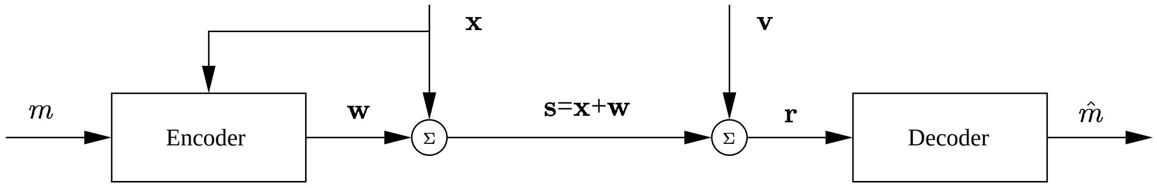


Fig. 1. Blind watermark communication facing an AWGN attack

Watermark communication as shown in Fig. 1 can be considered as *communication with side-information at the encoder.* This has been first realized in 1999 by Chen and Wornell [7] and Cox, Miller and McKellips [8]. Chen and Wornell introduced an important but almost forgotten paper by Costa into the watermarking community. Costa [9] showed theoretically that the channel capacity for the communication scenario depicted in Fig. 1 with an IID Gaussian host signal $\mathbf{x}$ is

$$C_{\mathrm{ICS}} = \frac{1}{2}\log_2\left(1 + \frac{\sigma_w^2}{\sigma_v^2}\right), \tag{1}$$

independent of σ_x^2. The suffix "ICS" stands for *ideal Costa scheme,* and is used here to distinguish the theoretical performance limit from that of suboptimal schemes discussed below. The performance of ICS depends solely on the watermark-to-noise power ratio $\mathrm{WNR} = 10\log_{10}(\sigma_w^2/\sigma_v^2)$ [dB]. The result (1) is surprising since it shows that the original data $\mathbf{x}$ need not be considered as interference at the decoder although the decoder does not know $\mathbf{x}$. Costa presents a theoretic scheme which involves a random codebook $\mathcal{U}^{L_x}$ which is

$$\begin{aligned}\mathcal{U}^{L_x} \quad &= \quad \{\mathbf{u}_l = \mathbf{w}_l + \alpha\mathbf{x}_l \;\mid\; l \in \{1, 2, \ldots, L_{\mathcal{U}}\}, \\ &\qquad \mathbf{w} \sim \mathcal{N}(0, \sigma_w^2 I_{L_x}), \mathbf{x} \sim \mathcal{N}(0, \sigma_x^2 I_{L_x})\}, \end{aligned} \tag{2}$$

where $\mathbf{w}$ and $\mathbf{x}$ are realizations of two L_x-dimensional independent random processes $\mathbf{w}$ and $\mathbf{x}$ with Gaussian probability density function (PDF). $L_{\mathcal{U}}$ is the total number of codebook entries and I_{L_x} denotes the L_x-dimensional identity matrix. For secure watermarking, the codebook choice must be dependent on a key K. There exists at least one such codebook such that for $L_x \to \infty$ the capacity (1) is achieved. Note that the optimum choice of the parameter α depends on the WNR and is given by

$$\alpha^*_{\mathrm{I}CS} = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_v^2} = \frac{1}{1 + 10^{-\mathrm{WNR[dB]}/10}}. \tag{3}$$

The ideal Costa scheme (ICS) is not practical due to the involved huge random codebook. Therefore, several suboptimal implementations of ICS have been proposed since 1999. A natural simplification of ICS is the usage of a structured codebook $\mathcal{U}^{L_x}$, which in the most simple case can be constructed by a concatenation of scalar uniform quantizers. This approach, constrained to a sample-wise (scalar) embedding and extraction rule, is denoted in this article as scalar Costa scheme (SCS). The accurate and complete performance analysis of SCS is the main topic of this paper.

Before discussing SCS, we give a brief review of related research on the implementation of Costa's scheme. Chen and Wornell developed in 1998 quantization index modulation (QIM) which provides good performance for low channel noise, but is not robust for channel conditions with $\sigma_v^2 > \sigma_w^2$ [10], [1]. In 1999, they improved the QIM idea using Costa's approach and named the new scheme *QIM with distortion compensation* (DC-QIM) [11]. Most of the work of Chen and Wornell concentrates on high dimensional embedding techniques where the dimensionality tends to infinity. This approach enables the analytical derivation of performance bounds. However, little is said about the performance of currently implementable schemes. Further, simulation results using state-of-the-art channel coding techniques are not provided. Chen and Wornell also discuss a simplification of DC-QIM where the indexed quantizers are derived via dithered prototype quantizes. This technique is investigated particularly for the case of uniform scalar prototype quantizers, which is denoted as distortion compensated dither modulaton (DC-DM). Chen and Wornell present a coarse performance analysis of DC-DM that is based on minimum-distance arguments and the variances of the watermark and the attack noise. However, the specific shape of the involved PDFs of the transmitted and received signals are not modelled accurately so that tight performance limits cannot be computed.

Ramkumar and Akansu [12], [13], [14], [15], [16] propose a blind watermarking technique based on periodic embedding and reception functions for self-noise suppression (host signal interference reduction). In particular low dimensional versions of this approach, e.g., with scalar embedding and reception functions, are closely related to suboptimal implementations of Costa's scheme. Ramkumar and Akansu consider during their analysis a proper modelling of the PDFs of the transmitted signals. However, their analysis of the receiver performance involves approximations that are only valid if adjacent codebook entries for identical messages are far from each other. This assumption is not valid for a large range of practically relevant WNRs. Further, Ramkumar and Akansu present a capacity analysis based on an equivalent noise variance derived from the PDFs of the transmitted signal. This analysis is a good approximation only for low WNRs. For high WNRs, evaluation of the presented capacity formula results in values above the Shannon limit. Nevertheless, it should be emphasized that certain versions of the technique proposed by Ramkumar and Akansu show good performance particularly for very low WNRs. SCS outperforms their approach in the range of typical WNRs only slightly [17].

Chou et al. [18] exploit the duality of communication with side-information at the encoder to source coding with side-information at the decoder to derive a watermarking scheme based on trellis-coded quantization. This work can be considered as an extension of the research on practical implementations of Costa's scheme in the direction of high dimensional embedding and reception rules. However, research in this direction is difficult and little progress has been made within the last years. Up to now, performance results that are better than the theoretical capacity limit of ST-SCS propose (see Sec. V) have not been published. Latest results by Chou et al. [19] show at least a slight improvement of turbo coded trellis-based constructions over simple SCS communication using coded modulation techniques. Note also that

SCS communication might still remain attractive due to its simplicity even if superior performance of high dimensional embedding techniques can be shown in future.

Note also that principles of Costa's work on communication with side information have recently gained some attention within multiuser communications [20], [21], [22].

The goal of this paper is to summarize theoretical and experimental results on the performance of the practical SCS embedding and reception technique. An accurate performance analysis is based on properly derived PDFs of the transmitted and received data. Further, a comparison of the performance of state-of-the art coding techniques with theoretical performance limits is given. In Sec. II, SCS is derived formally and the encoding and decoding process is outlined. Theoretical performance limits of SCS are derived in Sec. III. Experimental results for SCS communication at high rates (low noise power) are given in Sec. IV. Sec. V discusses SCS communication at low rates, which is particularly important for robust digital watermarking. Sec. VI discusses the important extension of the AWGN attack to an attack with additional amplitude scaling. An efficient algorithm for the estimation of such amplitude scaling attacks is presented. Finally, the invertibility of SCS watermark embedding is investigated in Sec. VII, which is of interest if the distortion introduced by watermark embedding should be reduced or even removed by the legal user of a watermarked document.

II. Scalar Costa Scheme

For a practical implementation of Costa's scheme, the usage of a suboptimal, structured codebook is proposed, while leaving the main concept of Costa's scheme unchanged. Besides being practical, the developed scheme is independent from the data distribution. This property can be achieved for a properly chosen embedding key sequence $\mathbf{k}$ [5], [6]. When no key is used, a reasonably smooth PDF $p_{\mathsf{x}}(x)$ and $\sigma_{\mathsf{x}}^2 \gg \sigma_{\mathsf{w}}^2, \sigma_{\mathsf{v}}^2$ must be assumed. To obtain a codebook with a simple structure, $\mathcal{U}^{L_x}$ is chosen to be a product codebook of dithered uniform scalar quantizers, which is equivalent to an L_x-dimensional cubic lattice [23].

A. SCS Encoder

First, the watermark message $m \equiv \mathbf{b}$, where $\mathbf{b}$ is a binary representation of m, is encoded into a sequence of watermark letters $\mathbf{d}$ of length L_x. The elements d_n belong to a D-ary alphabet $\mathcal{D} = \{0, 1, \ldots, D-1\}$. D-ary signaling denotes SCS watermarking with an alphabet $\mathcal{D}$ of size $D = |\mathcal{D}|$. In many practical cases, binary SCS watermarking ($d_n \in \mathcal{D} = \{0, 1\}$) will be used.

Second, the L_x-dimensional codebook $\mathcal{U}^{L_x}$ of Costa's scheme is structured as a product codebook

$\mathcal{U}^{L_x} = \mathcal{U}^1 \circ \mathcal{U}^1 \circ \cdots \circ \mathcal{U}^1$ of L_x one-dimensional component codebooks $\mathcal{U}^1$, where all component codebooks are identical. For D-ary signaling, the component codebook $\mathcal{U}^1$ must be separated into D distinct parts so that

$$\mathcal{U}^1 = \mathcal{U}^1_0 \cup \mathcal{U}^1_1 \cup \cdots \cup \mathcal{U}^1_d \cup \cdots \cup \mathcal{U}^1_{D-1}. \tag{4}$$

The codebook $\mathcal{U}^1$ is chosen to be equivalent to the representatives of a scalar uniform quantizer with step size $\alpha\Delta/D$, which is formally denoted as

$$\mathcal{U}^1 = \left\{ u = l\alpha\Delta + \frac{d}{D}\alpha\Delta \,\middle|\, d \in \mathcal{D}, l \in \mathbb{Z} \right\}. \tag{5}$$

l enumerates all quantizer representatives of a prototype scalar quantizer with step size $\alpha\Delta$, and d introduces a shift of the prototype quantizer. The dth sub-codebook of $\mathcal{U}^1$ is given by

$$\mathcal{U}^1_d = \left\{ u = l\alpha\Delta + \frac{d}{D}\alpha\Delta \,\middle|\, l \in \mathbb{Z} \right\}, \tag{6}$$

so that each sub-codebook is equivalent to the representatives of a scalar uniform quantizer with step size $\alpha\Delta$.

A simple and efficient encryption method for the SCS codebook is the derivation of a cryptographically secure pseudo-random sequence $\mathbf{k}$ from the watermark key K, with $k_n \in [0, 1)$, and the modification of each component codebook so that

$$\mathcal{U}^1(k_n) = \left\{ u_n = (l + k_n)\alpha\Delta + \frac{d_n}{D}\alpha\Delta \,\middle|\, d_n \in \mathcal{D}, l \in \mathbb{Z} \right\}. \tag{7}$$

Without knowing $\mathbf{k}$, it is practically impossible to reconstruct the codebook $\mathcal{U}^{L_x}(K)$ used for watermark embedding. Note that the presented encryption does not modify any codebook properties being important for communication reliability, e.g., the distance between different codebook entries.

For a Costa-type embedding of the watermark letters $\mathbf{d}$ a jointly typical pair $(\mathbf{u}_0, \mathbf{x})$ has to be found, which is equivalent to finding a sequence $\mathbf{q} = \mathbf{w}/\alpha = (\mathbf{u}_0/\alpha) - \mathbf{x}$ which is nearly orthogonal to $\mathbf{x}$. This search can be considered also as quantization of $\mathbf{x}$ with an L_x-dimensional quantizer, where each quantizer representative is derived from the codebook entries $\mathbf{u} \in \mathcal{U}^{L_x}$ via $\mathbf{u}/\alpha$. We propose a scheme in that this process is reduced to the sample-wise operation

$$\begin{aligned} q_n &= \mathcal{Q}_\Delta\left\{ x_n - \Delta\left(\frac{d_n}{D} + k_n\right)\right\} \\ &\quad - \left(x_n - \Delta\left(\frac{d_n}{D} + k_n\right)\right), \end{aligned} \tag{8}$$

where $\mathcal{Q}_\Delta\{\cdot\}$ denotes scalar uniform quantization with step size Δ. Finally, the transmitted watermark sequence is given by

$$\mathbf{w} = \mathbf{u}_0 - \alpha\mathbf{x} = \alpha\mathbf{q}, \tag{9}$$

and the watermarked data is

$$\mathbf{s} = \mathbf{x} + \mathbf{w} = \mathbf{x} + \alpha\mathbf{q}. \tag{10}$$

A block diagram of the presented watermark embedding scheme is depicted in Fig. 2. Fig. 3 shows an example input-output characteristic for $\alpha = 0.6$. The embedding of $d_n \in \mathcal{D}$ can be expressed as subtractive dithered quantization, where $\Delta(k_n + d_n/D)$ is the dither sequence and Δ is the step size of the uniform scalar quantizer. Note that the quantization error $\mathbf{q}$ and thus also $\mathbf{w}$, is almost orthogonal to the quantizer input $\mathbf{x}$ for an almost uniform original data PDF in the range of one quantization bin. For the given codebook encryption by a uniformly distributed key sequence $\mathbf{k}$, it can even be shown [24], [25] that $\mathbf{q}$ and $\mathbf{w}$ are statistically independent from $\mathbf{x}$, as it is in Costa's ideal scheme. Further, the power of the quantization error is always $\mathrm{E}\left\{q^2\right\} = \Delta^2/12$ for the given distribution of the key sequence.

SCS embedding depends on two parameters: the quantizer step size Δ and the scale factor α. For a given watermark power σ_w^2, these parameters are related by

$$\alpha = \sqrt{\frac{\sigma_w^2}{\mathrm{E}\left\{q^2\right\}}} = \sqrt{\frac{12\sigma_w^2}{\Delta^2}} = \frac{\sigma_w\sqrt{12}}{\Delta}. \tag{11}$$

Costa [9] determined α^*_{ICS}, as defined in (3), to be the optimum value of α for the codebook (2). For a suboptimal codebook, e.g., the product codebook of scalar uniform quantizers used in SCS, the optimum value of α can be different. However, no analytical solution has been found yet. In Sec. III-C, the optimum value for α in SCS depending on the WNR is computed numerically.

B. SCS Decoder

SCS decoding is very similar to the decoding process in ICS, except that the product codebook $\mathcal{U}^{L_x}(K) = \mathcal{U}^1(k_0) \circ \mathcal{U}^1(k_1) \circ \cdots \circ \mathcal{U}^1(k_{L_x-1})$, with $\mathcal{U}^1(k_n)$ as in (7), is used. Treating this codebook as a quantizer, the decoder acts as if it quantizes the received data $\mathbf{r} = \mathbf{x} + \mathbf{w} + \mathbf{v}$ to the closest codebook entry. From this view of the decoding process, a sound interpretation of the encoding process results: The encoder perturbs the original data $\mathbf{x}$ by $\mathbf{w}$ to form the transmitted data $\mathbf{s} = \mathbf{x} + \mathbf{w}$ so that, with high probability, $\mathbf{r}$ will fall into the correctly indexed quantization bin. Simple hard-decision decoding of the nth watermark letter d_n is achieved by scalar quantization of r_n with $\mathcal{U}^1(k_n)$.

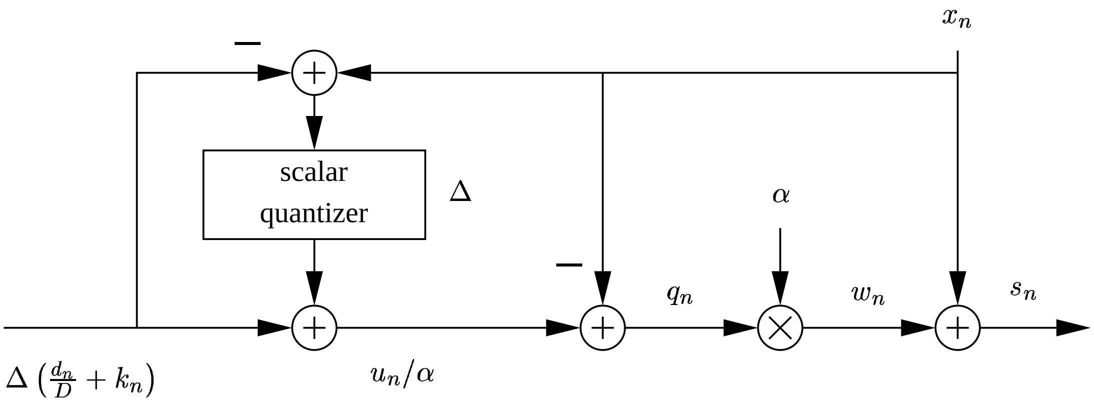


Fig. 2. SCS watermark embedding of the watermark letter d_n, encrypted with key k_n, into the original data element x_n.

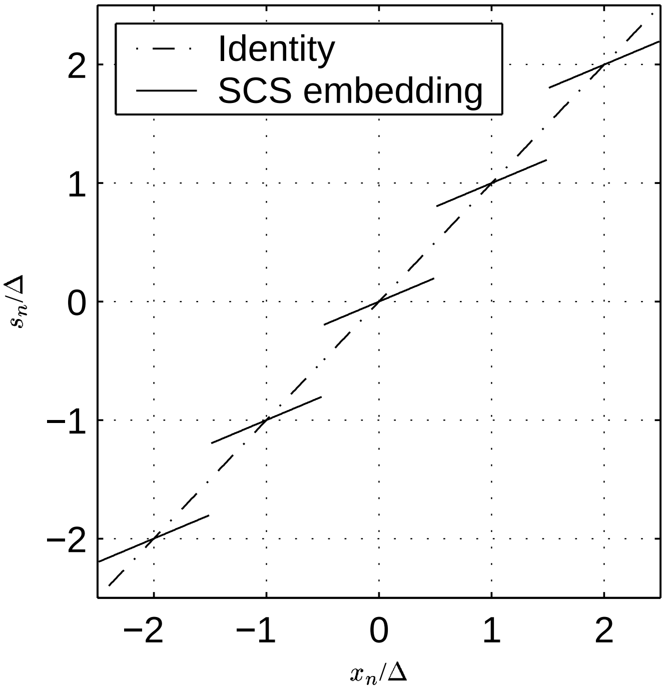


Fig. 3. Input-output characteristic for SCS embedding ($\alpha = 0.6$; $d_n = 0$; $k_n = 0$).

In general, the decoding reliability can be improved by decoding an entire watermark letter sequence $\hat{\mathbf{d}}$, where the known encoding of m into $\mathbf{d}$ can be exploited to estimate the most likely $\hat{\mathbf{d}}$, or equivalently, to estimate the most likely watermark message $\hat{m}$. The simple codebook structure of SCS can be exploited to efficiently estimate $\hat{\mathbf{d}}$. First, data $\mathbf{y}$ is extracted from the received data $\mathbf{r}$. This extraction process operates sample-wise, where the extraction rule for the nth element is

$$y_n = \mathcal{Q}_\Delta\left\{r_n - k_n\Delta\right\} - (r_n - k_n\Delta). \tag{12}$$

For binary SCS, $|y_n| \leq \Delta/2$, where y_n should be close to zero if $d_n = 0$ was transmitted, and close to $\pm\Delta/2$ for $d_n = 1$. Second, depending on the type of error correction encoding of $\mathbf{d}$, soft-input decoding algorithms, e.g. a Viterbi decoder for convolutional codes, can be used to decode from $\mathbf{y}$ the most likely transmitted watermark message $\hat{m}$.

C. Quantization Index Modulation and Dither Modulation

Quantization index modulation (QIM), as described in [10], [26] is a special case of Costa's scheme, where $\alpha = 1$ regardless of the WNR. As a result, QIM can achieve the capacity of ICS as the WNR tends to infinity. However, for $\sigma_v^2 > \sigma_w^2$ which is relevant in watermarking applications, reliable communication is difficult since the quantizer cells are too small. Dither modulation (DM) with scalar prototype quantizers, as described in [10], [26], relates to a general QIM scheme like SCS relates to an ideal Costa scheme. Note that throughout this paper DM is always considered to operate with uniform scalar prototype quantizers. Then, DM can be considered a special case of SCS, where $\alpha = 1$ regardless of the WNR. Since α is optimized in SCS for each WNR (see Sec. III-C), SCS can never perform worse than DM. In [11], [27], Chen and Wornell discuss the extension of QIM using Costa's ideas, and denote the derived scheme as QIM with distortion compensation which is basically Costa's scheme described in a different way.

III. Theoretical Performance Analysis

The performance loss of SCS compared to ICS, and a performance comparison of SCS and DM is of interest. Further, the optimization of the parameter α in SCS is desired. Performance is considered in terms of the watermark capacity of the specific schemes in case of an AWGN attack. The basis for an accurate performance evaluation are stochastic models for the watermarked data $\mathbf{s}$ and for the extracted data $\mathbf{y}$. With these stochastic models, the capacities of SCS and DM in case of AWGN attacks are computed. The capacity computation for SCS involves the optimization of the parameter α.

A. Distribution of Watermarked Data

Due to the simple codebook structure in SCS and DM, the sample-wise embedding and extraction procedure, and the IID original data, $\mathbf{s}$ can be considered a realization of an IID stochastic process s with the PDF $p_{\mathsf{s}}(s)$. For performance evaluation of the considered watermarking schemes, the conditional PDFs $p_{\mathsf{s}}(s|d, k)$ for all $d \in \mathcal{D}$ are required. Conditioning on the key k is necessary since otherwise the key hides any structure of the watermarked data. For simplicity, $k = 0$ is assumed for the presented illustrations.

SCS and DM are based on uniform scalar quantization with step size Δ. It is assumed that the host data is almost uniformly distributed over the range of several quantizer bins. This assumption is reasonable in most watermarking applications, where the host-data power is much stronger than the watermark power ($\sigma_x^2 \gg \sigma_w^2$). Note that the introduced assumptions may no longer be valid in case of SCS embedding for strong attacks since Δ might become quite large. For the following analysis it is not necessary to accurately model the PDF $p_s(s|d,k)$ for all possible values of s. It is sufficient to have an accurate model for s in the range of several quantizer bins. Thus, for mathematical convenience, $p_s(s|d,k)$ is considered periodic with period Δ.

With the introduced assumptions, the shape of one period of $p_s(s|d=0,k)$, denoted by $p_{\tilde{s}}(s|d=0,k)$, can be easily derived from the embedding rules for SCS and DM:

$$\begin{aligned} \text{DM: } p_{\tilde{s}}(s|d=0,k) &= \delta(s) && (13)\\ \text{SCS: } p_{\tilde{s}}(s|d=0,k) &= \frac{1}{\Delta(1-\alpha)}\text{rect}\left(\frac{s}{\Delta(1-\alpha)}\right). && (14)\end{aligned}$$

$\delta(\cdot)$ denotes the Dirac impulse and the rectangular signal is $\text{rect}(a) = 1$ for $|a| \leq 0.5$ and $\text{rect}(a) = 0$ for $|a| > 0.5$. The PDFs $p_s(s|d \neq 0,k)$ are almost identical to $p_s(s|d=0,k)$ except for a shift by $d/D \cdot \Delta$, that is

$$p_s(s|d,k) = p_s\left(s - \frac{d}{D}\Delta \middle| d=0,k\right) \qquad (15)$$

for both watermarking schemes.

Fig. 4 depicts qualitatively the PDFs of the transmitted value s in case of binary signaling ($d \in \mathcal{D} = \{0,1\}$) for both considered schemes. Note that $p_s(s|d=0,k)$ and $p_s(s|d=1,k)$ may even overlap for low WNR and that the choice of Δ can be quite different for both schemes which is not reflected in Fig. 4.

B. *Distribution of Extracted Received Data*

In this article, attacks by AWGN *independent* from the original data and the watermark signal are considered. Thus, the PDF of the received data r is given by the convolution of the PDF of the transmitted data s and the PDF $p_v(v)$ of the additive channel noise:

$$\begin{aligned} p_r(r|d,k) &= p_s(r|d,k) * p_v(r) && (16)\\ p_r(r|k) &= \frac{1}{|\mathcal{D}|}\sum_{d\in\mathcal{D}} p_r(r|d,k), && (17)\end{aligned}$$

where '$*$' denotes convolution. (17) is valid for $\text{Prob}(d) = 1/|\mathcal{D}|$. Since it is assumed that $p_s(s|d,k)$ is periodic with period Δ, $p_r(r|d,k)$ is also periodic with period Δ. One such period of $p_r(r|d,k)$

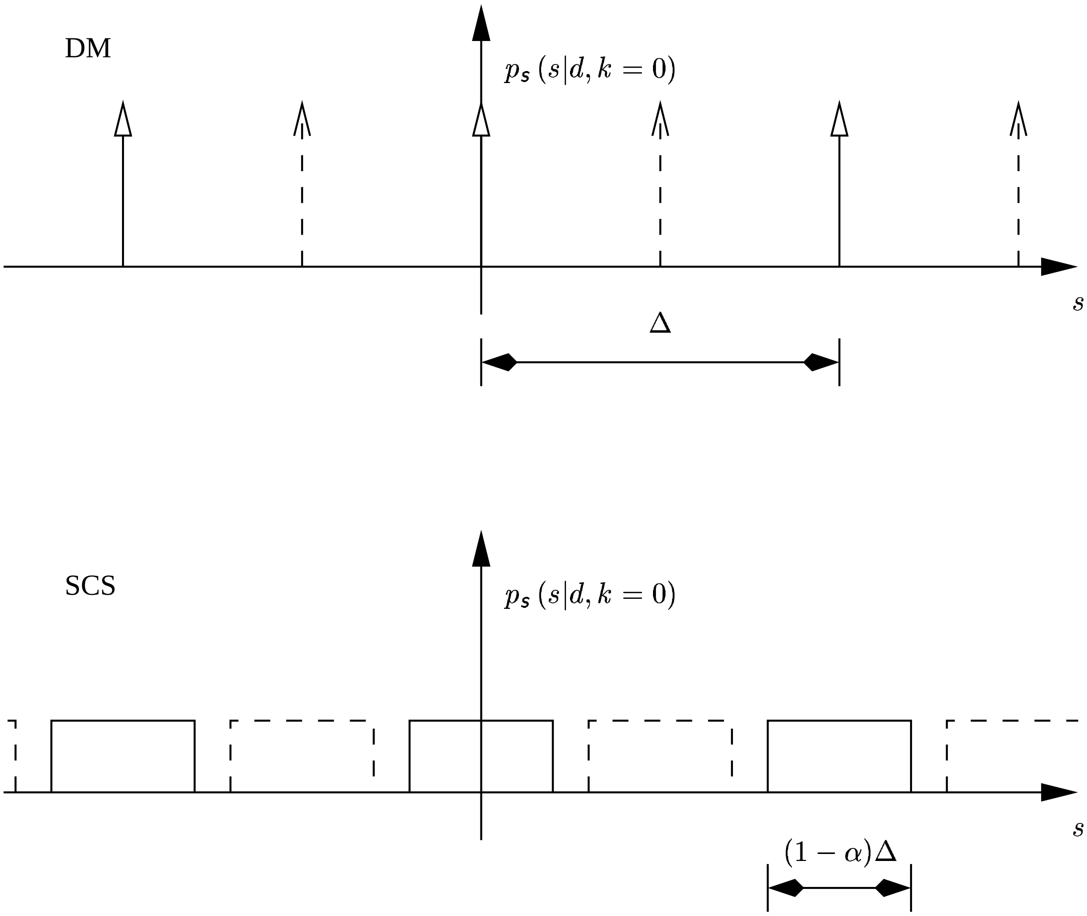


Fig. 4. Qualitative diagram of the PDFs $p_s\left(s|d=0,k=0\right)$ ("—") and $p_s\left(s|d=1,k=0\right)$ ("– –") of the watermarked data s for binary dither modulation (DM) and the scalar Costa scheme (SCS).

is identical (except for a normalization factor) to the PDF $p_y\left(y|d,k\right)$ of the extracted data y, where extraction with the correct key is assumed. A simple analytical expression for $p_y\left(y|d,k\right)$ is not known. Thus, $p_y\left(y|d,k\right)$ is computed numerically as described in [5], [6], [17]. Note that even for DM and a Gaussian PDF $p_v\left(v\right)$ the PDF of the received value will not be exactly Gaussian. Periodically overlapping Gaussian PDFs have to be considered due to the multiple representation of the watermark letters.

The upper plot of Fig. 5 depicts one period of the PDF of the watermarked elements s conditioned on the transmitted watermark letter d, and $k=0$ for binary SCS. The lower plot shows the respective PDFs of the extracted received elements y after an AWGN attack. In case of using an incorrect key $\mathbf{k}$ at the receiver, the distribution of $p_y\left(y|d\right)$ will be uniform for any possible received signal. This is indicated by the dotted line in the lower plot of Fig. 5.

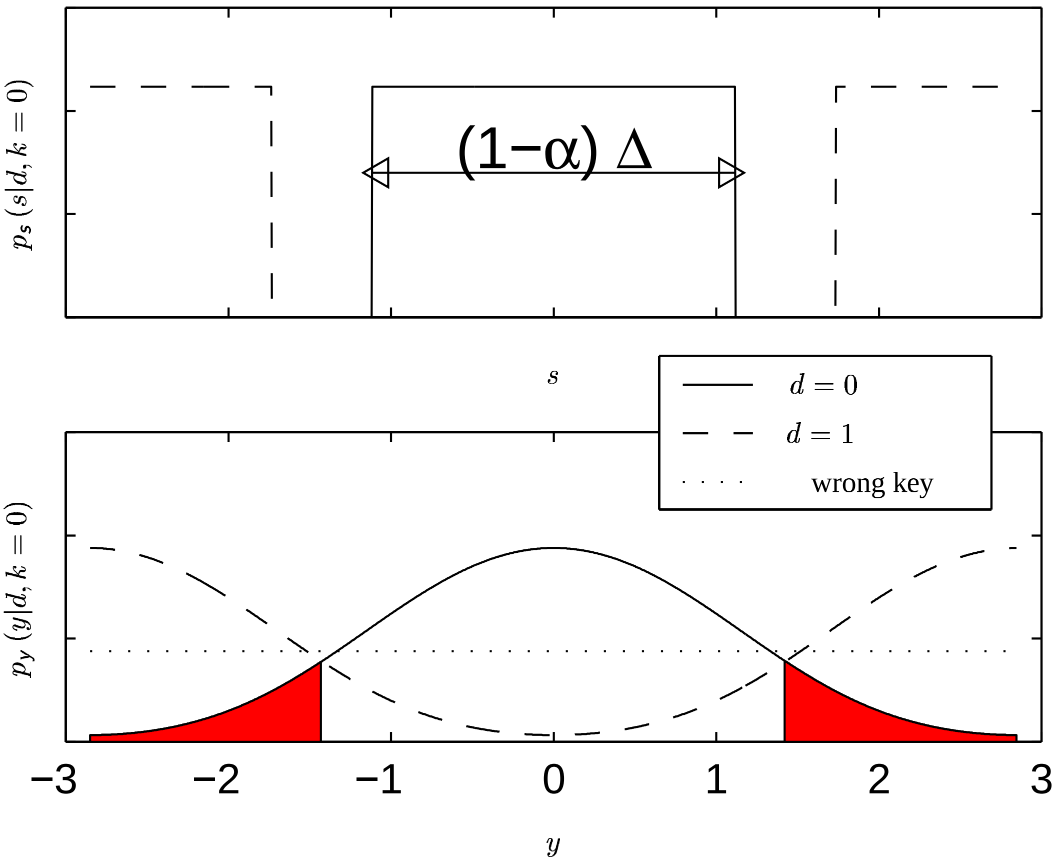


Fig. 5. One period of the PDFs of the watermarked data s and the extracted data y for binary SCS (σ_w^2=1, WNR $= 2$ dB, $\Delta = 5.7$, $\alpha = 0.61$). The filled areas represent the probability of decision errors assuming $d = 0$ was transmitted.

C. Capacity Computation and Optimum SCS Step Size

For the discrete memoryless channel, Gel'fand and Pinsker [28] and Heegard and El Gamal [29] showed that, for communication with side information at the encoder, the capacity is given by

$$C = \max_{p_{u;w}(u,w|x)} (I(u;r) - I(u;x)), \tag{18}$$

where the maximum is taken over all joint PDFs of the form $p_x(x)p_{u;w}(u,w|x)\ p_r(r|w,x)$ and where u is an auxiliary random variable and $I(u;r)$ and $I(u;x)$ denote the mutual information between u and the received data r and the mutual information between u and the side information (the original data) x, respectively. At the encoder, a specific realization of u is chosen depending on the message m to be transmitted and the side information $\mathbf{x}$ available to the encoder. Appropriate realizations of u for all possible messages m and all possible side information $\mathbf{x}$ are listed in a codebook $\mathcal{U}$, which must be known to the encoder and decoder. (18) shows that the capacity of communication with side information at the encoder is given by the difference of information that the codebook $\mathcal{U}$ gives about the received data $\mathbf{r}$ and about the side information $\mathbf{x}$. In general, maximization over all possible codebooks $\mathcal{U}$ and over all corresponding embedding functions is required. Here, the capacity of the suboptimum schemes DM and SCS is considered. SCS is constrained to a codebook based on scalar quantizers, which are parameterized by α and Δ as shown in (5). α and Δ are related for fixed embedding distortion by (11). Thus, there is only one free codebook parameter for fixed embedding distortion so that the capacity of SCS is given by

$$C_{\mathrm{SCS}} = \max_{\alpha} I(y; d). \tag{19}$$

DM is a special case of SCS with $\alpha = 1$ so that the capacity of DM is directly given by

$$C_{\mathrm{DM}} = I(y; d)\Big|_{\alpha=1}. \tag{20}$$

The watermark message is encoded such that for each data element an alphabet $\mathcal{D} = \{0, 1, \dots, D-1\}$ of watermark letters is used, where each letter is equiprobable. Then the mutual information $I(y; d)$ is given by [30]

$$\begin{aligned} I(y; d) \quad = \quad & -\int p_{y}\left(y|k\right) \log_2 p_{y}\left(y|k\right) \mathrm{d}y \\ & + \frac{1}{D} \sum_{d \in \mathcal{D}} \int p_{y}\left(y|d,k\right) \log_2 p_{y}\left(y|d,k\right) \mathrm{d}y. \end{aligned} \tag{21}$$

It can be observed that $I(y; d)$ is completely determined by the PDFs $p_{y}\left(y|k\right)$ and $p_{y}\left(y|d,k\right)$ as derived in Sec. III-B for the case of AWGN attacks.

For SCS, it is not possible to compute the maximization over α in (19) analytically since $p_{y}\left(y|k\right)$ and $p_{y}\left(y|d,k\right)$ are given numerically. Thus, α is optimized numerically for WNRs in the range of -20 dB to 20 dB. The resulting values for α are shown in Fig. 6 and corresponding values for Δ are shown in Fig. 7. An approximative analytical expression for the optimum value of α has been derived experimentally which is

$$\alpha_{\mathrm{SCS,\,approx}} = \sqrt{\frac{\sigma_{w}^2}{\sigma_{w}^2 + 2.71\sigma_{v}^2}}. \tag{22}$$

This leads with (11) to

$$\Delta_{\mathrm{SCS,\,approx}} = \sqrt{12\left(\sigma_{w}^2 + 2.71\sigma_{v}^2\right)}. \tag{23}$$

Fig. 6 and Fig. 7 show also the optimum value $\alpha^{*}_{\mathrm{ICS}}$ derived by Costa for ICS (see (3)) and the corresponding value for Δ when using $\alpha^{*}_{\mathrm{ICS}}$ in SCS. It can be observed that $\alpha^{*}_{\mathrm{ICS}}$ is almost identical to the

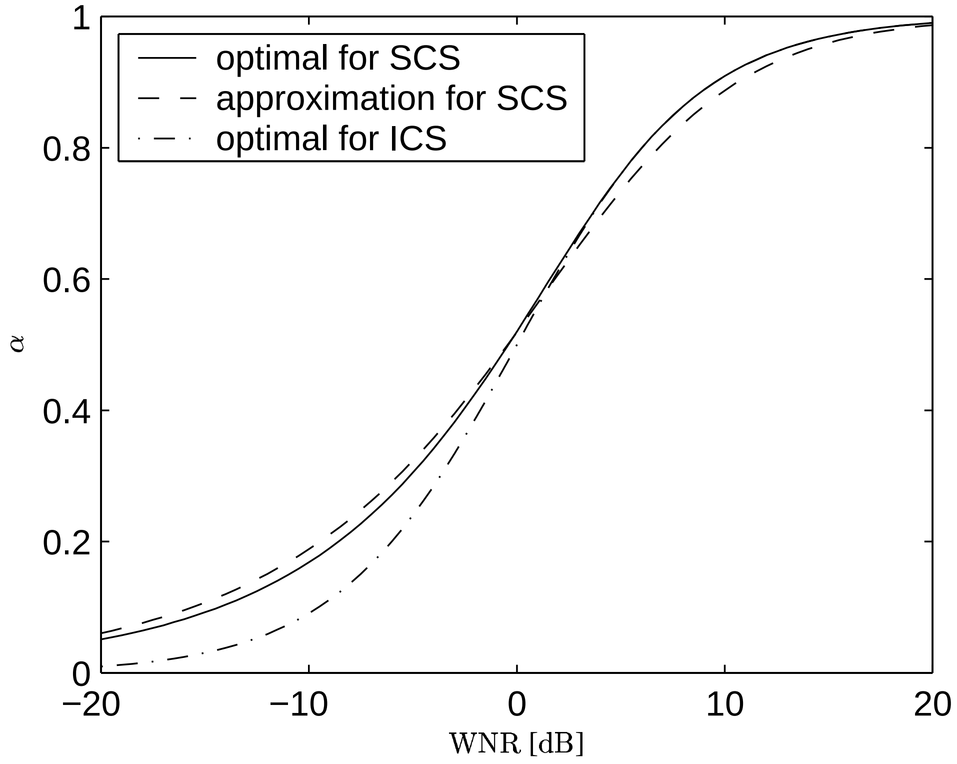


Fig. 6. Codebook parameter α

optimum value α^*_{SCS} for SCS in case of positive WNRs. However, for negative WNRs, α^*_{ICS} is too small for SCS. In this case, $\alpha_{\mathrm{SCS,\,approx}}$ defined in (22) is a better approximation for the optimum value of α for SCS.

D. Capacity of SCS Watermarking facing AWGN attacks

Fig. 8 compares the capacities of ICS, binary SCS, binary DM, and blind spread-spectrum (SS) watermarking for AWGN attacks. SCS watermarking does not achieve capacity, but is not too far from an ideal scheme either. DM performs poorly for negative WNRs, where the optimum value of α is significantly smaller than 1.

The term *spread-spectrum (SS) watermarking* has been established in the watermarking community for watermark embedding by the addition of a statistically independent pseudo-noise signal $\mathbf{w}$ with power σ_w^2 which is derived from the watermark message m and the key K. SS watermarking is one of the first methods used for watermarking (e.g., [31], [32]) and is still the most popular one. For a Gaussian original signal, AWGN attack, and for a Gaussian watermark signal $w \sim \mathcal{N}(0, \sigma_w^2)$ the capacity of blind

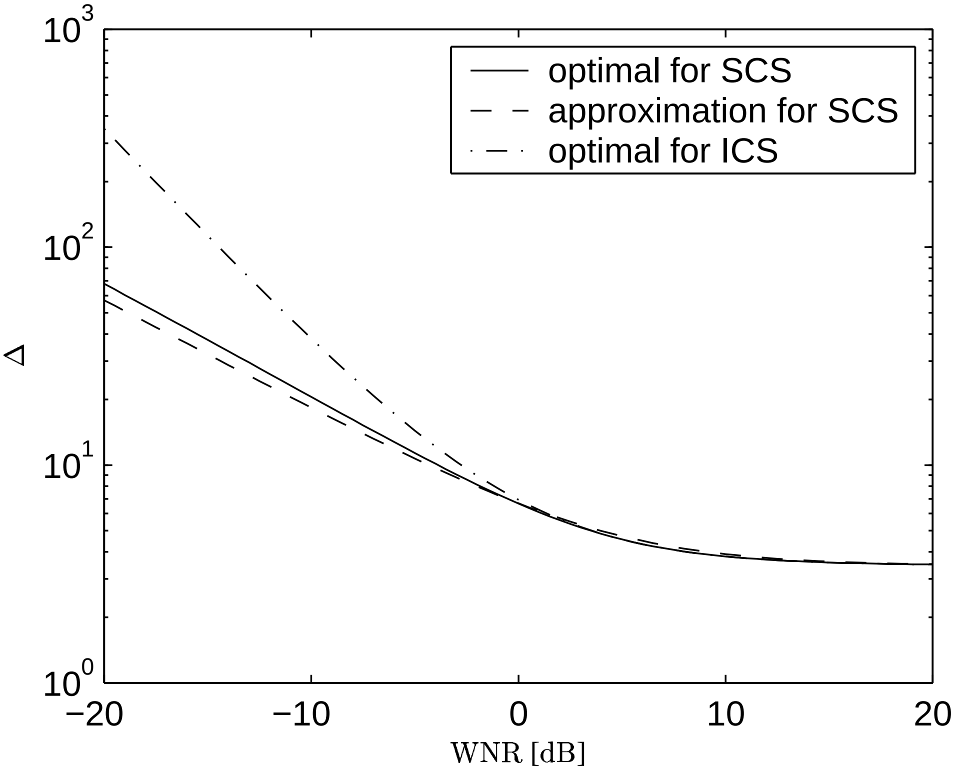


Fig. 7. Codebook parameter Δ for $\sigma_w^2 = 1$

SS watermarking is given by the capacity of an AWGN channel [30], that is

$$C_{\mathrm{SS}} = \frac{1}{2}\log_2\left(1 + \frac{\sigma_w^2}{\sigma_x^2 + \sigma_v^2}\right). \tag{24}$$

Note that $\sigma_x^2 \gg \sigma_w^2$ and $\sigma_x^2 \gg \sigma_v^2$ within common watermarking scenarios. Thus, the performance of blind SS watermarking facing an AWGN attack is mainly determined by the document-to-watermark power ratio $\mathrm{DWR} = 10\log_{10}(\sigma_x^2/\sigma_w^2)$ [dB]. This shows that blind watermark reception suffers significantly from original signal interference. The depicted capacity of blind SS watermarking is for $\mathrm{DWR} = 15$ dB. For weak to moderately strong attacks (i.e., WNRs greater than about -10 dB) SCS watermarking outperforms SS watermarking by far due to the data independent nature of SCS watermarking. However, Fig. 8 also reveals that for very strong attacks ($\mathrm{WNR} < -15$ dB), blind SS is more appropriate than SCS watermarking since here the attack distortion dominates possible interference from the original signal. Note that ICS outperforms blind SS watermarking for all WNRs.

Fig. 8 shows also that the binary SCS capacity is limited for high WNRs due to the binary alphabet $\mathcal{D}$ of watermark letters. Increasing the size D of the signaling alphabet $\mathcal{D}$ enables higher capacities for high WNRs as shown in Fig. 9. It can be observed that for very large signaling alphabets, the capacity

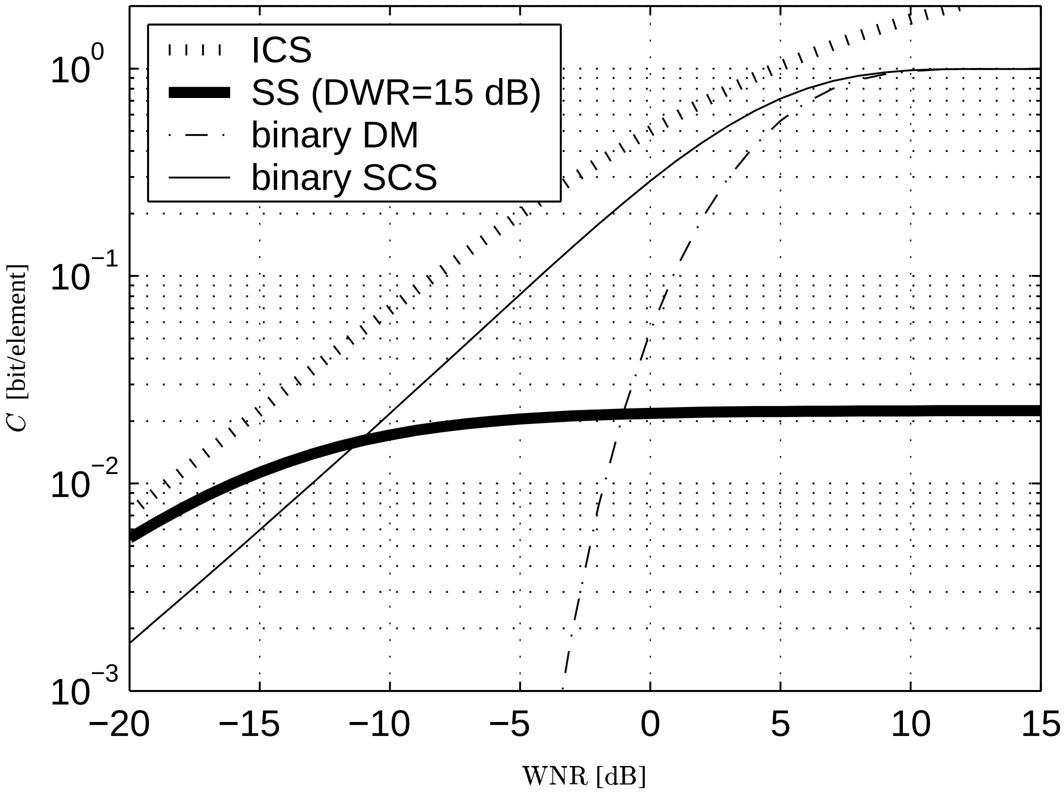


Fig. 8. Capacity of blind watermarking facing an AWGN attack compared for ICS, binary SCS, binary DM, and blind SS watermarking.

of SCS approaches slowly the capacity of ICS, or equivalently, $C_{\mathrm{SCS}}(\mathrm{WNR})$ is slightly steeper than $C_{\mathrm{ICS}}(\mathrm{WNR})$ for high WNRs.

IV. High-Rate SCS Communication

High-rate SCS communication is of interest for scenarios with low attack noise, for instance, if information embedding into analog channels is desired [33], [27], [34]. Here, information embedding at rates $R > 0.5$ bit/element is considered high-rate watermarking, since for these rates the capacity of binary SCS is significantly lower than for D-ary signaling with $D > 2$, as shown in Fig. 9. It can be observed that the size of the alphabet $\mathcal{D}$ has a significant influence only for WNRs larger than about ≈ 4 dB, or equivalently $R > 0.6$ bit/element.

Coded modulation techniques are used to combine D-ary signaling with binary error-correction coding. Here, the performance of SCS at $R = 1$ bit/element is investigated for different coded modulation techniques. As shown in Fig. 9, for $R = 1$ bit/element, 3-ary signaling is as good as D-ary signaling

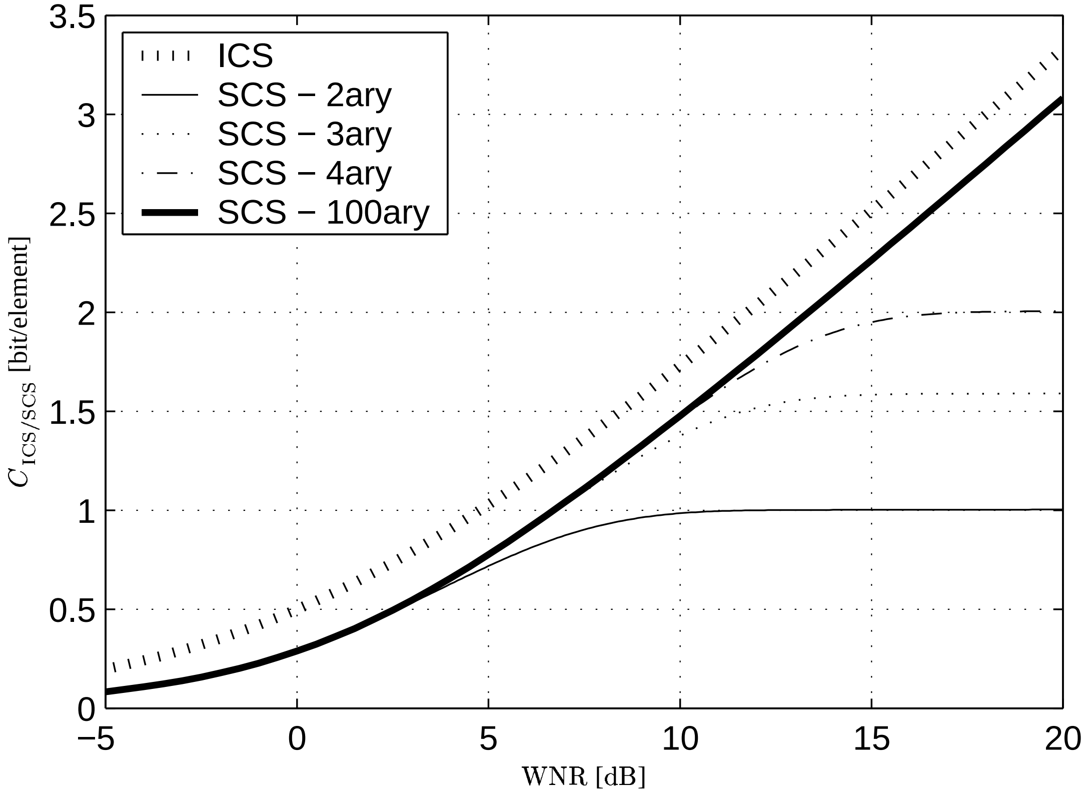


Fig. 9. Capacity of ICS and D-ary SCS watermarking facing an AWGN attack.

with $D > 3$. Examples for 4-ary and 8-ary signaling are discussed here due to their efficient combination with binary coding techniques. 4-ary signaling is used for the classical *trellis coded modulation with convolutional codes* (CC-TCM) as proposed by Ungerboeck [35]. 8-ary signaling is used in combination with a new *trellis coded modulation scheme with serial concatenated codes and iterative decoding* (SC-TCM). A detailed discussion of coded modulation is beyond the scope of this work. More details on these specific coded modulation schemes are given in [4], [5]. The main goal is to demonstrate that, with $R = 1$ bit/element, low bit-error rates (BER $< 10^{-5}$) can be achieved within 1.6 dB of the capacity C_{SCS}.

Bit-error rates (BERs) around 10^{-5} are achieved by CC-TCM and SC-TCM for WNR > 9.3 dB. The best performance for BER $< 10^{-5}$ was achieved by SC-TCM, with a minimum required WNR $\approx$ 8.3 dB. However, note that the computational complexity of ST-TCM and the codeword length (10,000 information bits) is also quite large compared to CC-TCM. Fig. 10 compares the measured minimum WNR for achieving BER $\leq 10^{-5}$ with SC-TCM and CC-TCM with the capacity of SCS and ICS. Ideally, SCS with $R = 1$ bit/element is possible for WNR > 6.7 dB. Thus, the discussed coded modulation

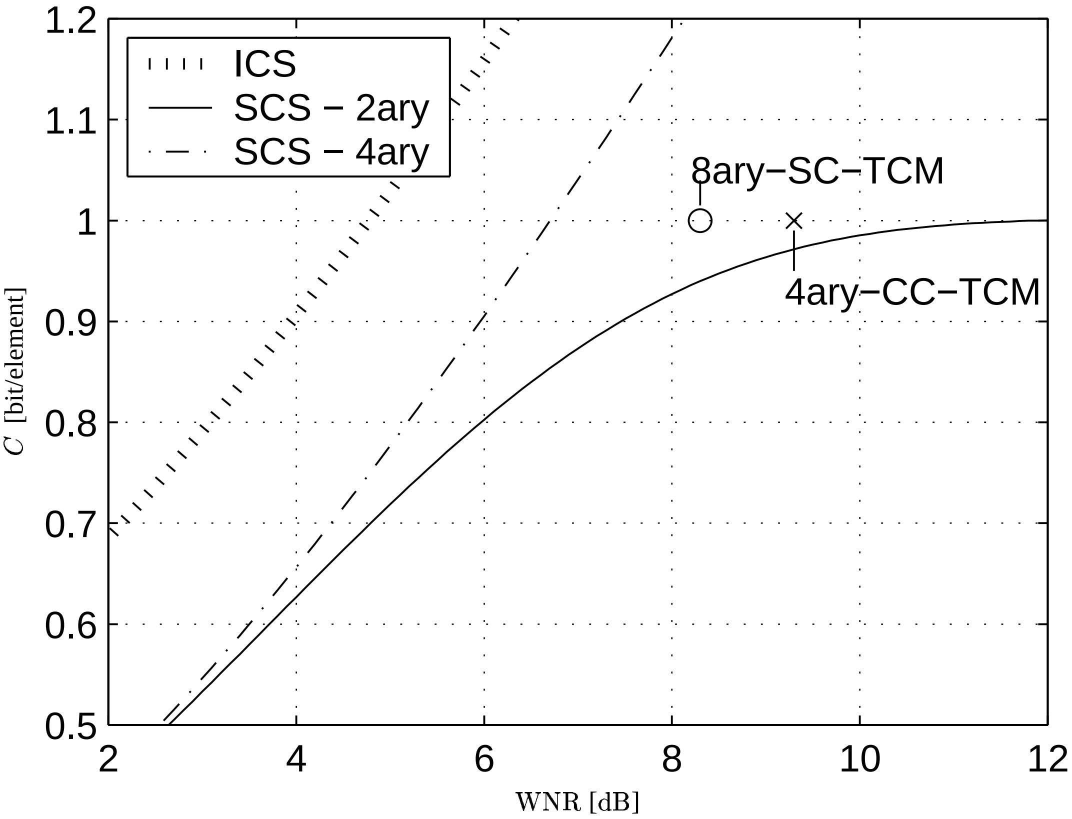


Fig. 10. SCS watermark capacity compared with measured results ($\mathrm{BER} \leq 10^{-5}$) using trellis coded modulation.

schemes come within $1.6 - 2.6$ dB of an optimal coding scheme for SCS. The distance to ICS is about 3.5 dB.

V. Low-Rate SCS Communication

In most watermarking applications, the attack distortion can be at least as large as the watermark embedding distortion. For the case of AWGN attacks, this means that a WNR of about 0 dB or less must be considered. For these distortion levels, reliable watermark communication can be achieved only at low rates.

Binary SCS is sufficient for low-rate watermarking. Thus, the watermark message m, represented by a binary sequence $\mathbf{b}$, has to be encoded into a sequence $\mathbf{b}_c = \mathbf{d}$ of binary watermark letters $d_n \in \{0, 1\}$. In order to achieve communication with low error rates, each bit of $\mathbf{b}$ has to be embedded redundantly into the original data $\mathbf{x}$. Different methods for the redundant embedding of $\mathbf{b}$ are investigated below, and their performance for an AWGN attack is compared.

A. Repetition Coding and Spread Transform

The simplest approach for the redundant embedding of the information bits $\mathbf{b}$ into the original data $\mathbf{x}$ is the repeated embedding of each bit. An alternative approach for redundant embedding of the information bits $\mathbf{b}$ into the original data $\mathbf{x}$ is the spread-transform (ST) technique as proposed in [7]. We found that repetition coding with SCS performs worse than ST with SCS (ST-SCS) which is not obvious at the first glance. Here, we illustrate the reason for this result.

Let ρ denote the *repetition factor* for SCS with repetition coding, e.g., one information bit is embedded into ρ consecutive data elements. However, instead of deciding for each extracted value y_n what transmitted watermark letter d_n is most likely, the decoder can directly estimate the most likely transmitted watermark information bit from ρ consecutive extracted values y_n [4], [5].

Spread transform watermarking has been proposed by Chen and Wornell [7]. A detailed description of this technique can be found in [7], [4], [5]. Here, we focus on the general principle. In ST watermarking, the watermark is not directly embedded into the original signal $\mathbf{x}$, but into the projection $\mathbf{x}^{\mathrm{ST}}$ of $\mathbf{x}$ onto a random sequence $\mathbf{t}$. Note that the term “transform”, as introduced by Chen and Wornell, is somewhat misleading since ST watermarking is mainly a pseudo-random selection of a signal component $\mathbf{x}^{\mathrm{ST}}$ to be watermarked. All signal components orthogonal to the spreading vector $\mathbf{t}$ remain unmodified. Let τ denote the *spreading factor*, meaning the number of consecutive original data elements x_n belonging to one element $x^{\mathrm{ST}}_{n'}$.

For watermark detection, the received data $\mathbf{r}$ is projected onto $\mathbf{t}$, too. The basic idea behind ST watermarking is that any component of the channel noise $\mathbf{v}$ that is orthogonal to the spreading vector $\mathbf{t}$ does not impair watermark detection. Thus, an attacker, not knowing the exact spreading direction $\mathbf{t}$, has to introduce much larger distortions to impair a ST watermark as strong as a watermark embedded directly into $\mathbf{x}$. For an AWGN attack, the effective WNR_τ after ST with spreading factor τ is given by

$$\mathrm{WNR}_\tau = \mathrm{WNR}_1 + 10\log_{10}\tau. \tag{25}$$

Thus, doubling the spreading length τ gives an additional power advantage of 3 dB for the watermark in the ST domain. However, note that repetition coding and ST with $\rho > 1$ and $\tau > 1$, respectively, achieve more robustness against attack noise at the cost of a reduced watermark rate. For a fair comparison of SCS with repetition coding and ST-SCS the watermark rate of both schemes should be equal, i.e. the repetition factor and the spreading factor should be equal ($\rho = \tau$).

The BERs for SCS with repetition coding and ST-SCS after an AWGN attack have been measured for different WNRs. Fig. 11 shows simulation results for $\rho = 2, 4, 8$ and $\tau = 2, 4, 8$. It has been observed that ST-SCS yields significantly lower BERs than SCS with repetition coding at the same watermarking rate. The predicted WNR gain of 3 dB for the same decoding reliability by doubling τ can be observed. However, the WNR gain for SCS with repetition coding is less than 3 dB when $\rho = 2$. The observed effect can be explained by examining the specific structure of the codebook $\mathcal{U}$ in SCS.

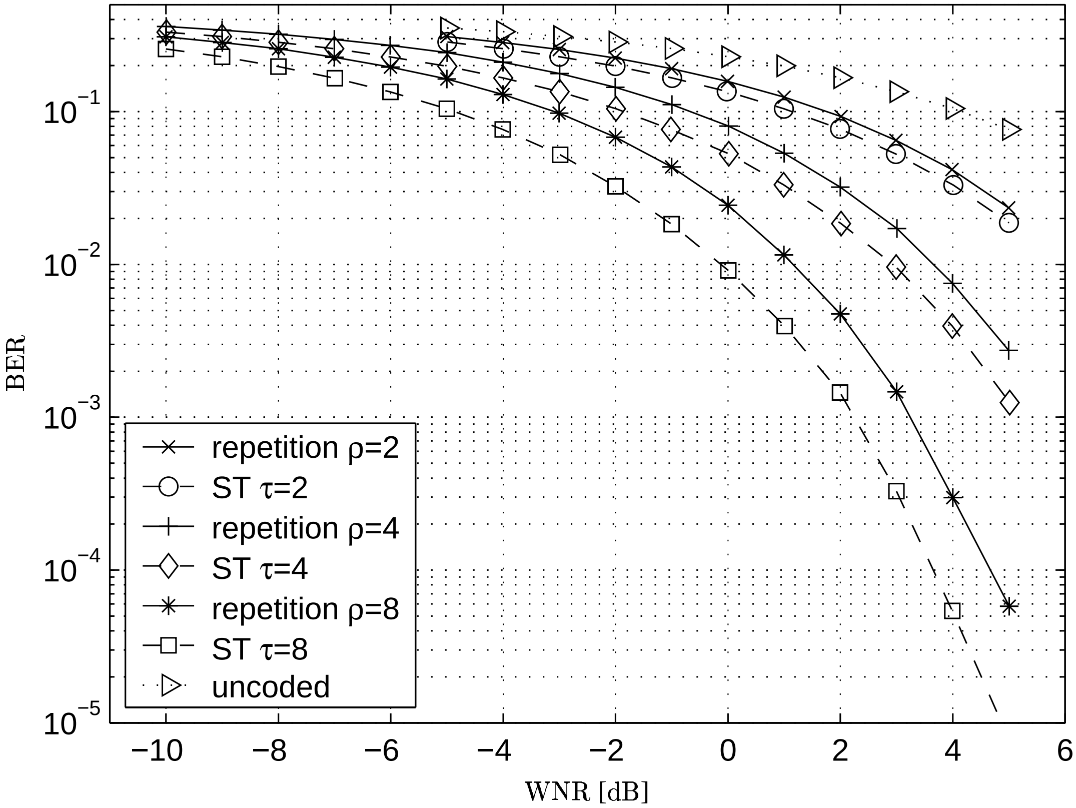


Fig. 11. BER for SCS with repetition coding and ST-SCS watermarking. For identical watermarking rates ($\tau = \rho$), ST-SCS yields lower error rates than SCS with repetition coding.

The multiple representations of a single watermark letter d_n by several points in the signaling space lead to many nearest neighbors which can lead to decoding errors. Fig. 12 shows a section of the two-dimensional PDFs of pairs of received data elements r_n in the case of an information bit $b_{n'} = 0$ for SCS with repetition coding with $\rho = 2$; bright areas indicate high probabilities. The key sequence $\mathbf{k}$ has been set to zero for illustration purposes. The circles and crosses depict the codebook entries corresponding to a transmitted watermark bit $b_{n'} = 0$ and $b_{n'} = 1$, respectively. Each circle is surrounded by four nearby crosses. Fig. 13 shows the corresponding two-dimensional PDFs in the case of ST-SCS with $\tau = 2$, where the spreading direction $\mathbf{t}$ was chosen to be the main diagonal. Obviously, any noise that is orthogonal to

$\mathbf{t}$ does not affect the decision whether the transmitted bit was 0 or 1. Further, each circle is surrounded only by two crosses. Thus, the probability that AWGN pushes watermarked data into the area where a decoding error occurs is lower for ST-SCS than for SCS with repetition coding.

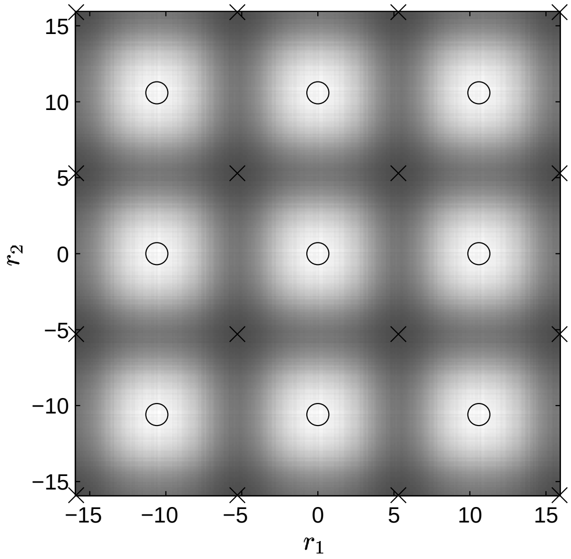


Fig. 12. Reception statistics for SCS with repetition coding with $\rho = 2$.

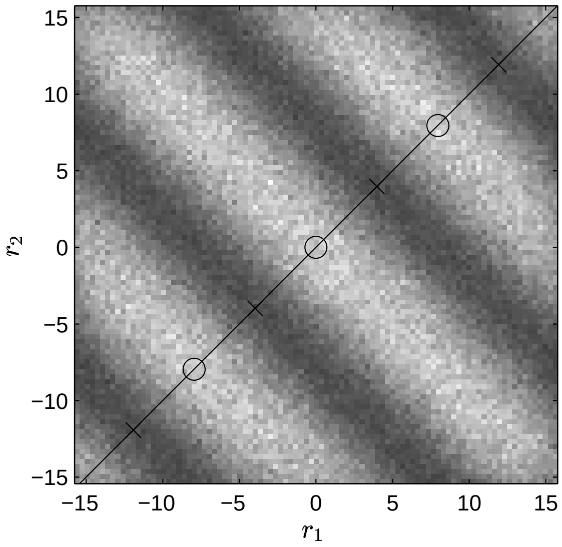


Fig. 13. Reception statistics for ST-SCS with $\tau = 2$.

Please note that the advantage of ST-SCS over SCS with repetition coding is only possible if the spreading direction $\mathbf{t}$ is not known to an attacker. Otherwise, an attacker would place all the noise in the direction $\mathbf{t}$ and the WNR-advantage vanishes. Further, ST-SCS watermarking with large spreading factors τ might be impractical since perfect synchronization of the complete spreading vector $\mathbf{t}$ is necessary. In contrast, decoding in the case of SCS with repetition coding is possible when only some of the watermarked data elements are synchronized. Another potential problem with large spreading factors τ is that the original-data power in the ST domain might become so low that the assumption that the original data is approximately uniformly distributed in the range of one quantizer cell no longer holds; this assumption is used in quantization based watermarking schemes like SCS and DM. As a consequence, the power of the watermark can no longer be predicted by $\Delta^2/12$. However, this problem can be avoided by using a key sequence $\mathbf{k}$ (Sec. II) that acts as a dither sequence that ensures a quantization noise power of $\Delta^2/12$.

B. Capacity of ST-Watermarking and Optimal Spreading Factor

ST-SCS watermarking should be considered a different suboptimal approach to implement a transmission scheme with side information at the encoder. Thus, the achievable rate of ST-SCS might be larger than that of SCS. Note that ST-SCS can never perform worse than SCS since SCS is a special case of ST-SCS with $\tau = 1$. The optimum choice of the spreading factor τ for attacks of differing noise powers is investigated.

Let $C_{\mathrm{ST},\tau}(\mathrm{WNR})$ denote the capacity of a specific watermarking scheme combined with a spread transform with spreading factor τ for an AWGN attack with given WNR. $C_{\mathrm{ST},1}(\mathrm{WNR})$ is the capacity of the respective scheme without ST. The performance of ST watermarking can be computed from that of the respective scheme without ST by

$$C_{\mathrm{ST},\tau}(\mathrm{WNR}) = \frac{C_{\mathrm{ST},1}(\mathrm{WNR}_\tau)}{\tau}, \tag{26}$$

with $\mathrm{WNR}_\tau = \mathrm{WNR} + 10\log_{10}\tau$.

Applying the ST technique the capacity of SCS and DM watermarking can be improved for WNRs lower than a certain $\mathrm{WNR}_{\mathrm{crit}}$ [4], [5]. For $\mathrm{WNR}s > \mathrm{WNR}_{\mathrm{crit}}$ the optimal spreading factor τ is 1, i.e. ST does not provide an additional capacity gain. The capacity of the ideal scheme ICS can not be improved by ST at all.

Fig. 14 shows the capacities of SCS, ST-SCS, DM, and ST-DM. Since the achievable rates for SCS and DM watermarking are computed numerically, the corresponding $\mathrm{WNR}_{\mathrm{crit}}$ are also obtained numerically. It can be found that for SCS, $\mathrm{WNR}_{\mathrm{crit,SCS}} \approx 0.01\mathrm{dB}$, and for DM, $\mathrm{WNR}_{\mathrm{crit,DM}} \approx 5.81$ dB. Fig. 14 shows also that DM can be improved significantly for $\mathrm{WNR} < \mathrm{WNR}_{\mathrm{crit,DM}}$, where for SCS only a minor gain is accessible. Note that ST-DM performs worse than simple SCS for most practical WNRs. Also, there is a constant gain of about 1.8 dB for ST-SCS over ST-DM in the range of negative WNRs.

C. SCS with State-of-the-Art Channel Coding

Repetition coding is known to be very inefficient. State-of-the-art error correction codes, e.g., turbo codes [36], outperform repetition coding by far. Therefore, simulation results for SCS communication using turbo coding are presented in Fig. 15.

Fig. 15 shows the minimum $\mathrm{WNR}_{\mathrm{min}}$ for which coded SCS watermarking gives BER $\approx 10^{-5}$. It can be observed that <u>t</u>urbo-<u>c</u>oded (TC) SCS performs indeed close to the capacity of SCS watermarking. The coding results for the code rates $R_\mathrm{c} = 1/2$ and $R_\mathrm{c} = 1/3$ can be translated to lower watermark rates R via ST watermarking, which is indicated by the straight lines. ST-SCS watermarking with a

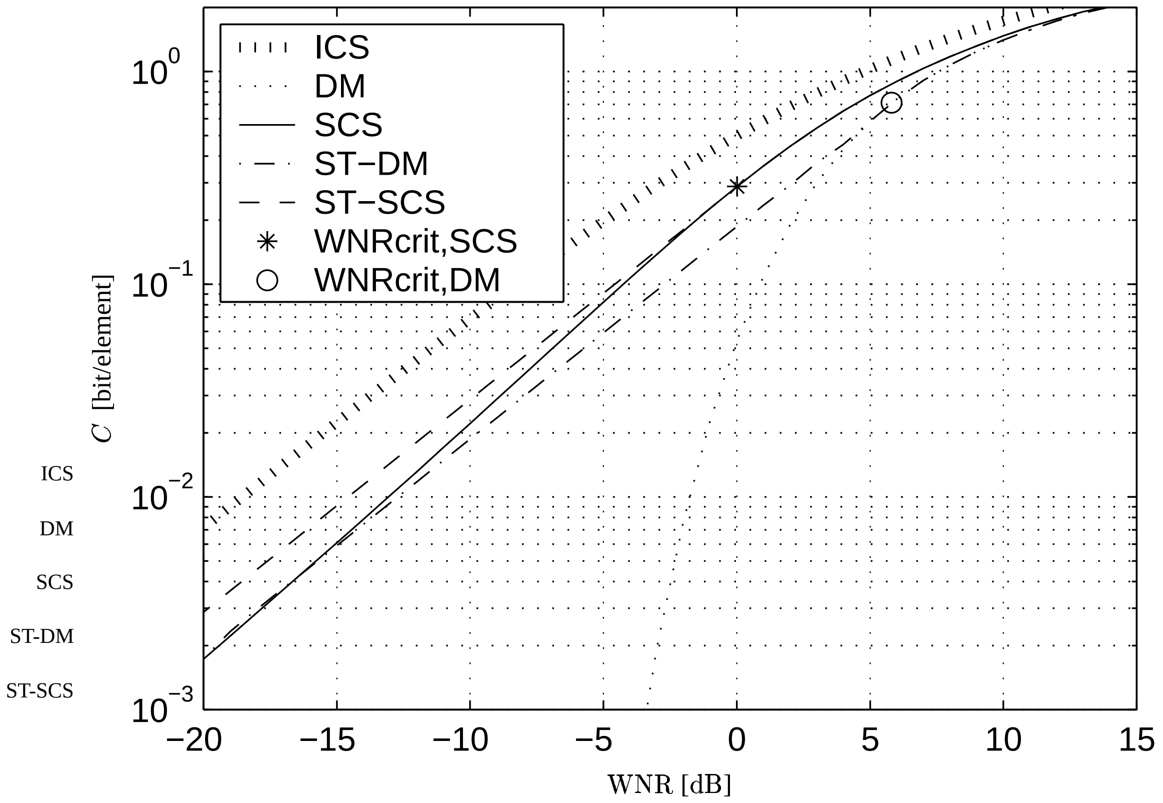


Fig. 14. Performance improvement by spread-transform watermarking.

code rate $R_\mathrm{c} = 1/3$ turbo code seems to be a very good choice for low-rate watermarking if any desired ST length τ is applicable. Fig. 15 shows also that turbo-coded SCS combined with repetition coding is less efficient than ST-SCS. Nevertheless, repetition coding might be useful in practice since it can be implemented in a very flexible way. Any received data element r_n with embedded watermark bit $b_{n'}$ increases the estimation reliability for $\hat{b}_{n'}$, where for ST watermarking all data elements r_n required for the computation of the projection $r^{\mathrm{ST}}_{n'}$ must be available to the receiver.

VI. SCS STEP SIZE ESTIMATION

A practically important extension of the simple AWGN channel model considered so far is a possible constant amplitude scaling and DC offset. Further, it might be of interest to adapt the SCS quantizer step size Δ to the characteristics of the original data. In both cases, a blind receiver is confronted with the difficulty of finding the proper quantizer step size for SCS reception. There exist several approaches to combat this problem. The step size Δ could be related to some statistics of the original that can be estimated robustly at the watermark decoder. A brute force approach would be to search for the valid step

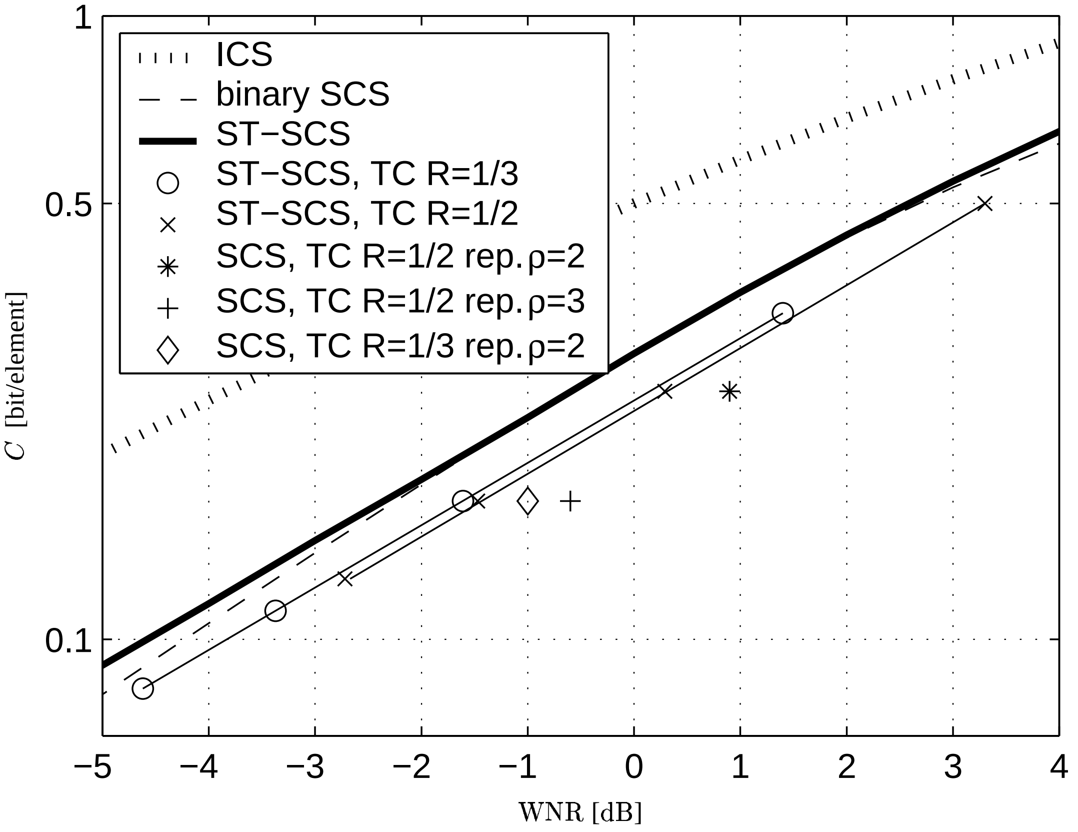


Fig. 15. Reliable coded binary (ST-)SCS compared with theoretical limits. The measured points indicate the minimum WNR for that a specific coding technique achieves BER $\approx 10^{-5}$.

size Δ by attempting to decode iteratively a valid message with different possible step sizes. Finally, we discuss here a specific step size estimation algorithm which is based on the analysis of histrograms of received data $\mathbf{r}$ with SCS watermarks [5], [37].

Fig. 16 depicts an extension of Fig. 1 where the attacker scales the watermarked data $\mathbf{s}$ by g (usually $g < 1$) and introduces additive white Gaussian noise (AWGN) $\mathbf{v}$, with $v \sim \mathcal{N}(r_{\text{offset}}, \sigma_v^2)$, that is

$$\mathbf{r} = g\mathbf{s} + \mathbf{v} = g(\mathbf{x} + \mathbf{w}) + \mathbf{v}. \tag{27}$$

Ideally, the receiver knows g and r_{offset} and thus compensates for the DC offset by subtracting r_{offset} and compensates for scaling by division by g (if $g \neq 0$). We characterize the attack strength by the effective watermark-to-noise power ratio $\text{WNR} = 10\log_{10}(g^2\sigma_w^2/\sigma_v^2)$ dB.

At the receiver, after compensation for g and r_{offset}, the extraction rule (12) can be applied to the signal $\mathbf{r}'$. However, if no compensation for g and r_{offset} is applied, the proper codebook for SCS watermark

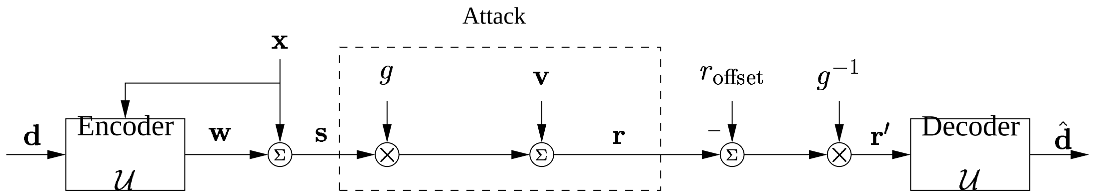


Fig. 16. Watermark communication facing an attack by amplitude scaling and AWGN with mean r_{offset}.

reception is

$$\hat{\mathcal{U}}_n^1(k_n) = \left\{ u_n = \left(l_n + \frac{d_n}{D} + k_n \right) \alpha\Delta_r + r_{\text{offset}} \middle| d_n \in \mathcal{D}, l_n \in \mathbb{Z} \right\}. \tag{28}$$

Here, $\Delta_r = g\Delta$ is the scaled quantizer step size which has to be used for SCS detection.

We propose a technique for estimating the attack channel parameters g and r_{offset} with the aid of a securely embedded pilot sequence $\mathbf{d}_{\text{pilot}} = \mathbf{0}$ of length L_{pilot}. Security is achieved again by embedding the pilot dependent on a secure random key sequence $\mathbf{k}$, where $k_n \in [0, 1)$. Note that estimation of $\Delta_r = g\Delta$ is sufficient to enable SCS watermark reception. g can be derived when Δ is known to the receiver.

The key idea behind our method for the estimation of Δ_r and r_{offset} is to analyze the 2-dimensional histograms of the received samples $r_{n,\text{pilot}}$ and the corresponding key values k_n, where $\mathbf{r}_{\text{pilot}} = (r_{0,\text{pilot}} \dots r_{n,\text{pilot}} \dots r_{L_{\text{pilot}}-1,\text{pilot}})$ is the sequence of received samples with embedded pilot symbols $\mathbf{d}_{\text{pilot}} = \mathbf{0}$. The suffix “pilot” is suppressed subsequently since only pilot samples are considered in this subsection.

Let $p_{r;k}(r, k)$ denote the 2-dimensional PDF of the received signal samples r_n and the corresponding key values k_n. Here, IID signals are considered so that the sample index n can be neglected in the statistical analysis.

Fig. 17 shows examples for $p_{r;k}(r, k)$, where incorrect keys and correct keys are assumed in the upper and lower plot, respectively. Note that for illustration purposes the step size Δ_r in this example is relatively large compared to the host signal standard deviation σ_x. Without knowing the correct $\mathbf{k}$, no structure in the watermarked signal is visible. The received PDF resembles basically the PDF of the host data which is a Gaussian distribution in the given example. However, computing $p_{r;k}(r, k)$ with the correct key k reveals the inherent structure in the data with embedded pilot samples. $p_{r;k}(r, k)$ shows stripes of high probability. For fix k, the distance between the peaks of two adjacent stripes equals the step size $\Delta_r = g\Delta$ (which gives $\Delta_r = 10$ in the shown example). This structure appears since the relevant quantizer structure within SCS embedding is dithered by the product $k\Delta$.

In practice, $p_{r;k}(r, k)$ has to be estimated via a 2-dimensional histogram, which requires proper dis-

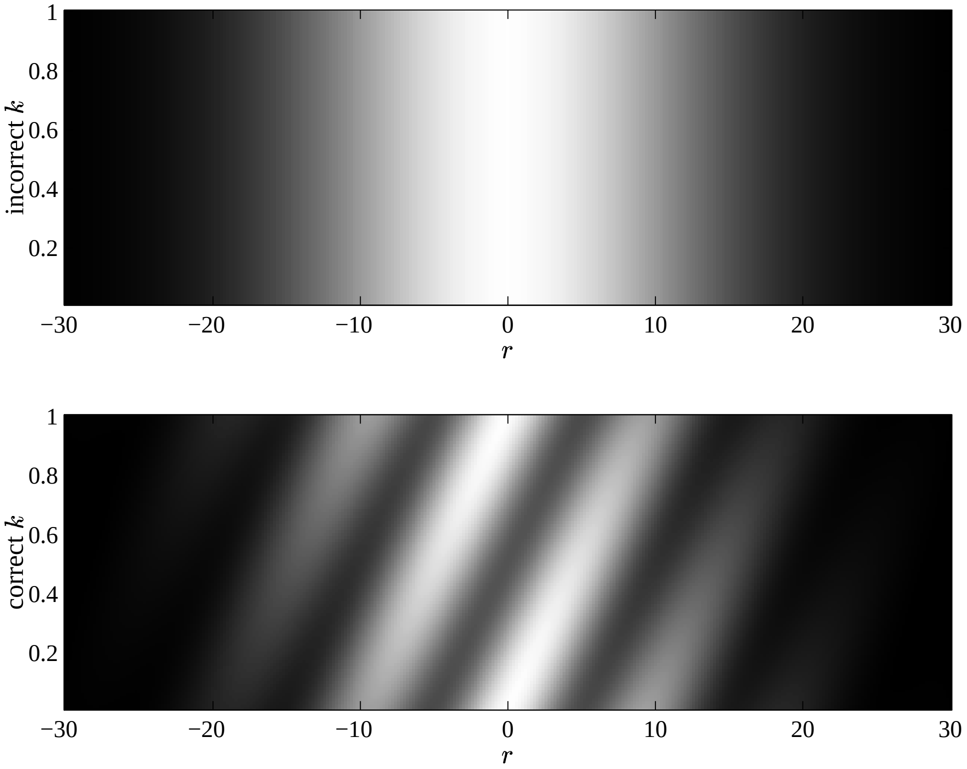


Fig. 17. 2-dimensional PDFs $p_{r;k}(r,k)$, where the upper plot is valid for reception with an incorrect key and the lower plot is valid for reception with the correct key. Bright areas indicate high probability.

cretization of the PDF variables (r,k). We describe here an estimation algorithm based on $L_{\mathrm{bin}k}$ different bins considered for the key value k. The one-dimensional conditional histograms of the samples r_n with key $k_n \in \mathbb{K}_v$ are analyzed separately, where

$$\mathbb{K}_v = \left\{ k \,\middle|\, \frac{v}{L_{\mathrm{bin}k}} \le k < \frac{v+1}{L_{\mathrm{bin}k}} \right\} \quad \text{for} \quad v \in \{0,1,\dots,L_{\mathrm{bin}k}-1\} \text{ and } L_{\mathrm{bin}k} > 1. \tag{29}$$

The $L_{\mathrm{bin}k}$ conditional histograms will show local maxima with a relative distance of Δ_r. The absolute position of these maxima gives an estimate of r_{offset}.

A. Parameter Estimation Based on Fourier Analysis

We introduce a simple model for the conditional PDFs $p_r(r|k \in \mathbb{K}_v)$ of the received pilot elements in order to motivate the afterwards described estimation of r_{offset}. The model is motivated by the observation that each PDF $p_r(r|k \in \mathbb{K}_v)$ shows local maxima with a distance of Δ_r. Let $p_r(r)$ denote the PDF of the received signal samples r_n. It can be assumed that $p_r(r)$ reflects more or less the host signal PDF ($p_r(r) \approx p_x(x)$) if the embedding distortion and attack distortion is small relative to the host signal power. An exact characterization of $p_r\left(r|k \in \mathbb{K}_\upsilon\right)$ is not necessary for our purpose. A sufficiently accurate model is given by

$$p_r\left(r|k \in \mathbb{K}_\upsilon\right) \approx \tilde{p}_r\left(r|k \in \mathbb{K}_\upsilon\right) = p_r\left(r\right)\left(1 + \gamma\cos\left(2\pi f_0 r - \Phi_0 - \frac{2\pi}{L_{\mathrm{bin}k}}\left(\upsilon + \frac{1}{2}\right)\right)\right), \quad (30)$$

where γ is an appropriate constant with $0 < \gamma < 1$. The model parameters f_0 and Φ_0 are directly related to the unknown parameters Δ_r and r_{offset}. f_0 determines the distance between two local maxima, and Φ_0 determines their absolute position. The exact relationship is given by

$$f_0 = \frac{1}{\Delta_r} \quad \text{and} \quad \Phi_0 = \frac{2\pi}{\Delta_r} r_{\mathrm{offset}} = 2\pi f_0 r_{\mathrm{offset}}. \quad (31)$$

Fig. 18 depicts an example for the given model. The local maxima of the conditional PDFs $\tilde{p}_r\left(r|k \in \mathbb{K}_\upsilon\right)$ with a relative distance of $\Delta_r = 10$ are clearly visible.

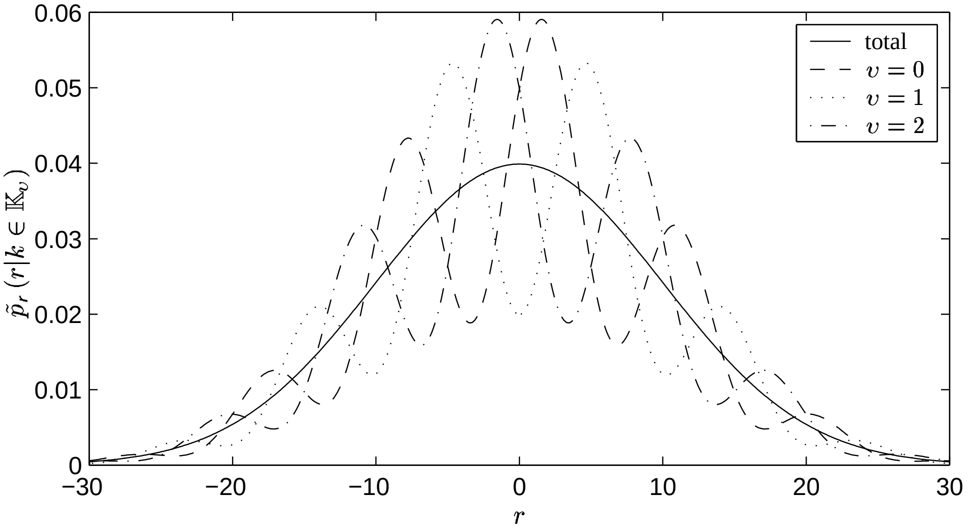


Fig. 18. Total and conditional PDFs of the received pilot sequence. $L_{\mathrm{bin}k} = 3$ different ranges for the key are distinguished. The example is for a Gaussian distribution of r_n, and for the parameters $\Delta_r = 10$ and $r_{\mathrm{offset}} = 0$.

The parameters f_0 and Φ_0 of the model given in (30) have to be computed from the given conditional PDFs $p_r\left(r|k \in \mathbb{K}_\upsilon\right)$ and the given unconditional PDF $p_r\left(r\right)$. Fourier analysis is appropriate for this task since f_0 and Φ_0 are the frequency and a constant phase contribution of the cosine term in (30).

For the υth conditional PDF, the normalized spectrum $A_{\upsilon(f)}$ is defined as

$$\begin{aligned} A_\upsilon(f) &= \mathcal{F}\left\{\frac{\tilde{p}_r\left(r|k \in \mathbb{K}_\upsilon\right)}{p_r\left(r\right)} - 1\right\} = \mathcal{F}\left\{\gamma\cos\left(2\pi f_0 r - \Phi_0 - \frac{2\pi}{L_{\mathrm{bin}k}}\left(\upsilon + \frac{1}{2}\right)\right)\right\} \\ &= \frac{\gamma}{2}\left[\mathrm{e}^{\mathrm{j}\left(-\Phi_0 - \frac{2\pi}{L_{\mathrm{bin}k}}\left(\upsilon + \frac{1}{2}\right)\right)}\delta\left(f_0 - f\right) + \mathrm{e}^{-\mathrm{j}\left(-\Phi_0 - \frac{2\pi}{L_{\mathrm{bin}k}}\left(\upsilon + \frac{1}{2}\right)\right)}\delta\left(f_0 + f\right)\right]. \quad (32) \end{aligned}$$

All $L_{\mathrm{bin}k}$ spectra can be combined in an elegant way due to the systematically different phase at $A_v(f_0)$ and $A_v(-f_0)$. The $L_{\mathrm{bin}k}$ spectra $A_v(f)$ are multiplied by $\mathrm{e}^{\mathrm{j}\frac{2\pi}{L_{\mathrm{bin}k}}v}$ prior to their summation to an overall spectrum $A(f)$, that is

$$A(f) \;=\; \sum_{v=0}^{L_{\mathrm{bin}k}-1} A_v(f)\,\mathrm{e}^{\mathrm{j}\frac{2\pi}{L_{\mathrm{bin}k}}v} = \frac{\gamma L_{\mathrm{bin}k}}{2}\,\mathrm{e}^{-\mathrm{j}\left(\Phi_0+\frac{\pi}{L_{\mathrm{bin}k}}\right)}\,\delta\left(f_0-f\right). \tag{33}$$

Thus, for the model given in (30), $|A(f)|$ has only one peak, which is located exactly at the frequency f_0. Further, $\Phi_0 = -\arg\{A(f_0)\} - \frac{\pi}{L_{\mathrm{bin}k}}$. Note that the multiplication by $\mathrm{e}^{\mathrm{j}\frac{2\pi}{L_{\mathrm{bin}k}}v}$ is superior to a multiplication by $\mathrm{e}^{\mathrm{j}\frac{2\pi}{L_{\mathrm{bin}k}}vf}$ which would correspond to a shift of the different conditional PDFs by $\frac{v}{L_{\mathrm{bin}k}}$. In the latter case, the spectrum $|A(f)|$ would have another peak at $f = -f_0$ which increases the required sampling interval for the numerical computation of the conditional PDFs.

The exact PDFs of the received signal do not fit exactly to the model given in (30). Further, in practice, the PDFs $p_r\left(r|k \in \mathbb{K}_v\right)$ and $p_r\left(r\right)$ can be only estimated from the L_{pilot} pilot samples $\mathbf{r}$. This estimation is obtained from histograms with $L_{\mathrm{bin}r}$ bins that cover the total range of all received samples. Based on these histograms, $A_v(f)$ is computed at $L_{\mathrm{DFT}} \geq L_{\mathrm{bin}r}$ discrete frequencies via a length-L_{DFT} DFT. Here, a single peak in the spectrum $A(f)$ cannot be exptected due to estimation errors and the inaccuracy of the model (30). Nevertheless, for L_{pilot} sufficiently large, a dominating peak should occur at f_0. Details of the outlined implementation are described in [5], [37].

B. *Estimation performance for different L_{pilot}*

The outlined algorithm for the estimation of Δ_r and r_{offset} is dependent on the following set of parameters:

L_{pilot}	:	length of pilot sequence
$L_{\mathrm{bin}r}$	:	number of histogram bins used for the pilot sample value
$L_{\mathrm{bin}k}$	:	number of histogram bins used for the key value
L_{DFT}	:	DFT length

The estimation accuracy also depends on the WNR, and on the DWR. In this paper, the estimation performance for different pilot length L_{pilot} is discussed for $\mathrm{WNR} = -10\ \mathrm{dB}, \ldots, 5\ \mathrm{dB}$. This range for WNR covers the most interesting range of attack strengths for that SCS watermarking might be useful. The DWR has been fixed to $\mathrm{DWR} = 20\ \mathrm{dB}$ and the remaining parameters are $L_{\mathrm{bin}r} = 50$, $L_{\mathrm{DFT}} = 1024$, and $L_{\mathrm{bin}k} = 5$. Experimental results that support this choice of parameters are given in [5].

The influence of the number L_{pilot} of received pilot elements is studied experimentally. For simplicity, $g = 1$ and no offset has been considered so that the estimator should ideally find $\hat{\Delta}_r = \Delta_r = \Delta$ and $\hat{r}_{\mathrm{offset}} = 0$. For the evaluation of the estimation performance, three different figures of merit have been used:

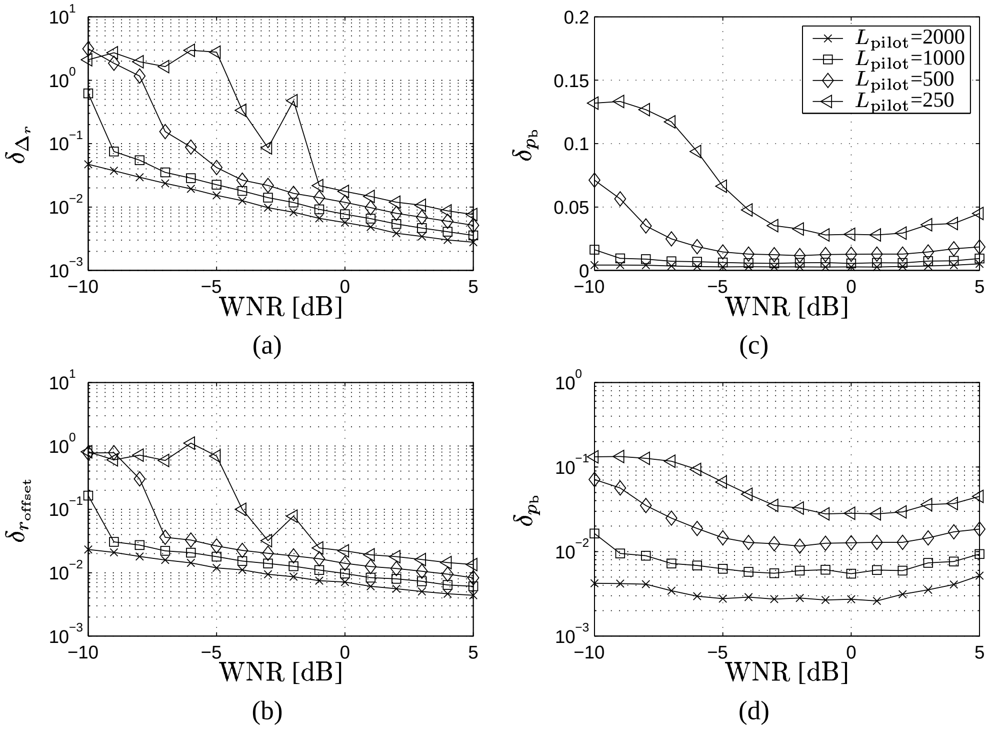
$$\text{relative error of } \hat{\Delta}_r: \quad \delta_{\Delta_r} = \frac{\sqrt{\mathrm{E}\left\{(\hat{\Delta}_r - \Delta_r)^2\right\}}}{\Delta_r}, \tag{34}$$

$$\text{relative error of } r_{\mathrm{offset}}: \quad \delta_{r_{\mathrm{offset}}} = \frac{\sqrt{\mathrm{E}\left\{(\hat{r}_{\mathrm{offset}} - 0)^2\right\}}}{\Delta_r}, \tag{35}$$

$$\text{relative increase of bit-error probability}: \quad \delta_{p_{\mathrm{b}}} = \frac{\mathrm{E}\left\{\hat{p}_{\mathrm{b}} - p_{\mathrm{b}}\right\}}{p_{\mathrm{b}}}. \tag{36}$$

δ_{Δ_r} and $\delta_{r_{\mathrm{offset}}}$ effectively measure the root of the mean squared estimation error relative to the exact step size Δ. These figures of merit have been chosen since not only the variance of estimation errors is important, but also a possible biased estimate. The relative increase of the bit-error probability p_{b} for uncoded binary SCS reception with estimated Δ_r and r_{offset} is given by $\delta_{p_{\mathrm{b}}}$. It is sufficient to measure the expected difference of the bit-error probability since imperfect estimates Δ_r and r_{offset} can only increase the bit-error probability on average. p_{b} of uncoded binary SCS is relatively high for the considered WNRs. However, many new parameters would have to be introduced for simulations with coded SCS communication, which would make a fair comparison more difficult. Further, the increase of p_{b} can be considered a good indicator for the effect of estimation errors on coded communication. The free parameters can be optimized only for a certain range of different WNRs where here the focus is on $\mathrm{WNR} = -5$ dB to $\mathrm{WNR} = 0$ dB. In particular the relative increase of the uncoded error probability $\delta_{p_{\mathrm{b}}}$ shows a local minimum for a certain WNR, since for large negative WNRs, the estimation accuracy is decreased due to the strong noise, and for high WNRs, the absolute decoding error is so low that any decoding error increases the relative decoding error significantly.

In general, it is desired to make the pilot sequence as short as possible, however, very short pilot sequences lead to an inaccurate PDF estimation, and thus to incorrect estimations of Δ_r and r_{offset}. Fig. 19 shows the estimation performance for $L_{\mathrm{pilot}} = 250, 500, 1000$, and 2000. Fig. 19.(a) depicts δ_{Δ_r} which describes the relative estimation error of Δ_r. For $L_{\mathrm{pilot}} = 2000$, δ_{Δ_r} decreases monotonically with increasing WNR, and is lower than 1% for $\mathrm{WNR} > -3$ dB. Shorter pilot sequences lead to an increased relative estimation error. However, for some WNR, robust estimation is no longer possible at all. Lowering the WNR further introduces so much noise into the PDF estimation that the largest component of the computed DFT spectrum appears at any random frequency index $0 < l < L_{\mathrm{DFT}} - 1$.



(a)

(c)

(b)

(d)

Fig. 19. Estimation performance for different pilot lengths L_{pilot} (DWR $=20$ dB, $L_{\text{bin}r}=50$, $L_{\text{bin}k}=5$).

For $L_{\text{pilot}}=250$, this effect occurs for $\text{WNR}<-1$ dB. For $L_{\text{pilot}}=500$, a minimum WNR of about -5 dB is required. Fig. 19.(b) depicts $\delta_{r_{\text{offset}}}$ which follows in general the behavior of δ_{Δ_r}. The resulting relative increase of the uncoded error rate δ_{p_b} is shown with linear and logarithmic axes in Fig. 19.(c) and Fig. 19.(d), respectively. δ_{p_b} increases monotonically with decreasing pilot length L_{pilot}. Further, it can be observed again that for some low WNR the estimation algorithm starts to fail completely. Nevertheless, it is quite promising that even for $L_{\text{pilot}}=500$, δ_{p_b} is lower than 2% for all $\text{WNR}\geq -5$ dB.

C. Estimation Based on SS Pilot Sequences

So far, an estimation of the SCS receiver parameter Δ_r based on a known SCS watermark has been proposed. However, it is also possible to estimate the scale factor g, and thus $\Delta_r=g\Delta$, with help of an additive spread-spectrum (SS) pilot watermark. Here, we present an analysis of the estimation accuracy δ_{Δ_r}, as defined in Sec. VI-B, when using SS pilot watermarks and compare the result with those for SCS pilot watermarks.

We consider again the attack channel defined in (27). However, now, we assume that $\mathbf{w}$ is a pseudo-noise sequence of length $L_w=L_{\text{pilot}}$ with zero mean ($\sum_{n=1}^{L_{\text{pilot}}} w_n=0$) and power $\sigma_w^2=\frac{1}{L_{\text{pilot}}}\sum_{n=1}^{L_{\text{pilot}}} w_n^2$. Throughout this analysis, an IID host signal $\mathbf{x}$ and additive noise signal $\mathbf{v}$ is assumed so that $x_n=x$ and

$v_n = v$, respectively. $\mathbf{w}$ is known to the watermark receiver so that g can be estimated from $\mathbf{r}$ based on the correlation $\hat{c}$ between $\mathbf{r}$ and $\mathbf{w}$, that is

$$\hat{c} = \frac{1}{L_{\text{pilot}}} \sum_{n=1}^{L_{\text{pilot}}} r_n w_n. \tag{37}$$

The unbiased estimate $\hat{g}$ of g derived from $\hat{c}$ is derived as follows:

$$\begin{aligned}
\mathrm{E}\{r_n\} &= \mathrm{E}\{g(x_n + w_n) + v_n\} = g\mathrm{E}\{x\} + \mathrm{E}\{v\} + g w_n && (38)\\
\mathrm{E}\{\hat{c}\} &= \frac{1}{L_{\text{pilot}}} \sum_{n=1}^{L_{\text{pilot}}} \mathrm{E}\{r_n\} w_n = \frac{1}{L_{\text{pilot}}} \sum_{n=1}^{L_{\text{pilot}}} (g\mathrm{E}\{x\} + \mathrm{E}\{v\}) w_n + \frac{g}{L_{\text{pilot}}} \sum_{n=1}^{L_{\text{pilot}}} w_n^2 \\
&= \frac{g\mathrm{E}\{x\} + \mathrm{E}\{v\}}{L_{\text{pilot}}} \underbrace{\sum_{n=1}^{L_{\text{pilot}}} w_n}_{=0} + \frac{g}{L_{\text{pilot}}} \sum_{n=1}^{L_{\text{pilot}}} w_n^2 = \frac{g}{L_{\text{pilot}}} \sum_{n=1}^{L_{\text{pilot}}} w_n^2 = g\sigma_w^2 && (39)\\
g &= \frac{\mathrm{E}\{\hat{c}\}}{\sigma_w^2} && (40)\\
\hat{g} &= \frac{\hat{c}}{\sigma_w^2} && (41)
\end{aligned}$$

(41) describes the estimation rule for $\hat{g}$ using the SS pilot watermark $\mathbf{w}$ that is known to the receiver. Next, the variance of $\hat{g}$ dependent on the pilot length L_{pilot} is derived. For simplicity, we assume that the host signal $\mathbf{x}$ and the attack noise $\mathbf{v}$ are mean-free ($\mathrm{E}\{x\} = 0$, and $\mathrm{E}\{v\} = 0$) so that the variance of $\mathbf{x}$ and $\mathbf{w}$ is given by $\sigma_x^2 = \mathrm{E}\{x^2\}$ and $\sigma_v^2 = \mathrm{E}\{v^2\}$, respectively. The derivation of the variance $\mathrm{Var}\{\hat{g}\}$ is tedious but not difficult so that only the main steps are presented here:

$$\begin{aligned}
\mathrm{E}\{\hat{c}^2\} &= (g\sigma_w^2)^2 + \frac{g^2\sigma_x^2 + \sigma_v^2}{L_{\text{pilot}}}\sigma_w^2 && (42)\\
\mathrm{Var}\{\hat{c}\} &= \mathrm{E}\{\hat{c}^2\} - \mathrm{E}\{\hat{c}\}^2 = \frac{g^2\sigma_x^2 + \sigma_v^2}{L_{\text{pilot}}}\sigma_w^2 && (43)\\
\mathrm{Var}\{\hat{g}\} &= \mathrm{Var}\left\{\frac{\hat{c}}{\sigma_w^2}\right\} = \frac{\mathrm{Var}\{\hat{c}\}}{(\sigma_w^2)^2} = \frac{g^2\sigma_x^2 + \sigma_v^2}{L_{\text{pilot}}\sigma_w^2} = \frac{g^2\sigma_x^2/\sigma_w^2 + \sigma_v^2/\sigma_w^2}{L_{\text{pilot}}}. && (44)
\end{aligned}$$

We observe that $\mathrm{Var}\{\hat{g}\}$ depends on the WNR via σ_x^2/σ_w^2 and on the DWR via σ_v^2/σ_w^2. The term σ_x^2/σ_w^2 dominates for realistic DWRs about 20 dB and WNR > -10dB. Further, we observe that $\mathrm{Var}\{\hat{g}\}$ decreases with increasing pilot length L_{pilot}.

Fig. 20 compares the achieved estimation accuracy using SS pilot watermarks and SCS pilot watermarks for $L_{\text{pilot}} = 1000$ and $L_{\text{pilot}} = 2000$. Note that the estimation accuracy δ_{Δ_r} for SS pilot watermarks can be computed theoretically from $\mathrm{Var}\{\hat{g}\}$ via

$$\delta_{\Delta_r} = \frac{\sqrt{\mathrm{E}\left\{(\hat{\Delta}_r - g\Delta)^2\right\}}}{g\Delta} = \frac{\sqrt{\mathrm{E}\{(\hat{g}\Delta - g\Delta)^2\}}}{g\Delta} = \frac{\sqrt{\mathrm{E}\{(\hat{g} - g)^2\}}}{g} = \frac{\sqrt{\mathrm{Var}\{\hat{g}\}}}{g}. \tag{45}$$

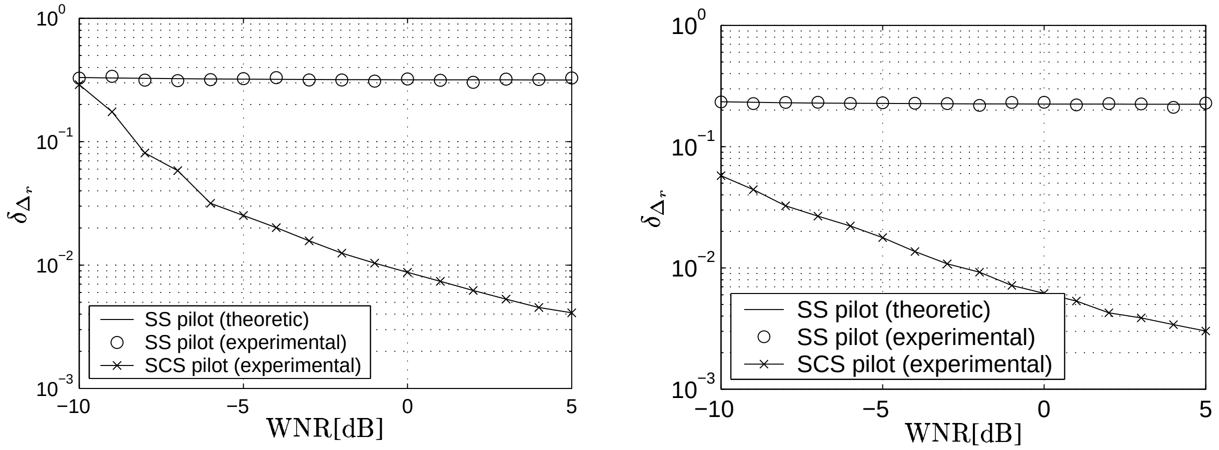


Fig. 20. Estimation performance for L_{pilot}=1000 (left) and L_{pilot}=2000 (right). The performance for SS pilot watermarks and SCS pilot watermarks is compared (DWR $= 20$ dB, $L_{\text{bin}k} = 3$, $L_{\text{bin}r} = 50$). The experimental results are averaged over 1000 simulations.

The results shown in Fig. 20 clearly demonstrate the superiority of the estimation algorithm based on SCS pilot watermarks. The advantage of the SCS pilot watermarks stems from the reduced influence of host-signal interference on the estimation accuracy.

VII. Inverse SCS

In some applications, it is desired to recover the original signal from the watermarked signal after watermark reception. Examples are information hiding applications dealing with medical images [38], or multiple watermark reception. In applications dealing with medical images, the goal is mainly to recover the original signal with a minimum amount of distortion. In multiple watermark reception, the interference of the first decoded watermark on other embedded watermarks should be minimized. For this, the already decoded watermark is exploited to remove the corresponding embedding distortion as much as possible.

Perfect recovery of the original signal might be impossible in many practical cases, e.g., attack noise cannot be removed in general. However, in some cases it is sufficient to produce a signal that is closer to the original signal than the received signal. In this section, ways to invert SCS watermarking are discussed. In practice, the receiver sees an attacked watermarked signal. Here, a simple AWGN attack is considered again. For completeness the noiseless case is discussed first. Throughout the section, it is assumed that the transmitted sequence of watermark letters $\mathbf{d}$ and the correct key sequence $\mathbf{k}$ are perfectly known, e.g., correct decoding has been performed, which can be treated without loss of generality as

$\mathbf{d} = \mathbf{0}$ and $\mathbf{k} = \mathbf{0}$. The effect of possible remaining bit errors after error correction decoding, and thus imperfect knowledge of $\mathbf{d}$, is not investigated. However, it is obvious that for low bit-error rates the influence of the incorrect inverse mapping applied to those samples with incorrectly received dither samples $\hat{d}_n$ on the overall quality improvement by inverse SCS is negligible.

A. Inverse SCS in the Noiseless Case

In the noiseless case, the watermark decoder receives the signal $\mathbf{r} = \mathbf{s}$. In this case, the deterministic embedding procedure can be inverted perfectly. With y_n from (12) the host signal x_n can be reconstructed with the next valid SCS codebook entry

$$r_{q,n} = \mathcal{Q}_\Delta\left\{r_n - \Delta\left(\frac{d_n}{D} + k_n\right)\right\} + \Delta\left(\frac{d_n}{D} + k_n\right) \tag{46}$$

by

$$x_n = r_{q,n} - \frac{y_n}{1-\alpha}. \tag{47}$$

The perfect invertibility of SCS is as well illustrated by the input-output characteristic of SCS embedding for $\alpha = 0.6$, $d_n = 0$, and $k_n = 0$ shown in Fig. 3. The input-output characteristic of SCS embedding is a strictly increasing function so that the inverse mapping in the noiseless case exists. This inverse mapping is obtained by mirroring the input-output characteristic of SCS embedding at that for the identity mapping $x_n = s_n$.

B. Inverse SCS after AWGN Attack

Inversion of SCS watermarking after transmission over an AWGN channel is considered. Contrary to the noiseless case, it is impossible to reconstruct $\mathbf{s}$ from the received signal $\mathbf{r}$ even with perfect knowledge of $\mathbf{d}$ because the transmitted value s_n depends also on the original signal value x_n that is not known to the receiver. Consequently, it is impossible to recover the host signal perfectly, however, one can at least try to find an estimate $\hat{\mathbf{x}}$ so that for the distortion holds $\mathrm{D}\left(\mathbf{x}, \hat{\mathbf{x}}\right) \leq \mathrm{D}\left(\mathbf{x}, \mathbf{r}\right)$, where the MSE distortion measure is adopted. In the following, it is assumed that the channel noise σ_v^2 is smaller than or equal to the noise variance $\sigma_{v,\mathrm{design}}^2$ for which the SCS watermark has been designed.

B.1 Estimation of the Original Signal

The minimum mean-squared error (MMSE) estimate $\hat{x}_n$ of the original signal sample x_n should be derived for each received sample r_n. IID signals are assumed so that the sample index n is suppressed in the following. With help of the known key sequence sample k and known watermark letter d, the deviation $y \in [-\frac{\Delta}{2}, \frac{\Delta}{2})$ from the next valid SCS codebook entry r_q is given by

$$y = r_q - r, \tag{48}$$

For AWGN attacks, the most likely corresponding quantized original signal sample is r_q. Thus, the MMSE estimate $\hat{x}$ is

$$\begin{aligned}\hat{x}(r_q, y) &= \underset{x_t \in \mathbb{R}}{\operatorname{argmin}}\ \mathrm{E}_{p_{\mathsf{x}}(x|r)}\left\{(x_t - \mathsf{x})^2\right\} \\ &= r_q - \underset{\hat{q}_t \in [-\frac{\Delta}{2}, \frac{\Delta}{2})}{\operatorname{argmin}}\ \mathrm{E}_{p_{\mathsf{x}}(x|r)}\left\{(r_q - \hat{q}_t - \mathsf{x})^2\right\} \\ &= r_q - \hat{q}(r_q, y),\end{aligned} \tag{49}$$

where r_q is no longer considered within the minimization, and $\hat{q}(r_q, y) \in [-\frac{\Delta}{2}, \frac{\Delta}{2})$ has to be chosen such that the MSE $\mathrm{E}_{p_{\mathsf{x}}(x|r)}\left\{(\hat{x} - \mathsf{x})^2\right\}$ is minimized. Straightforward analysis shows that $\hat{q}(r_q, y)$ has to be computed by

$$\hat{q}(r_q, y) = \mathrm{E}_{p_{\mathsf{x}}(x|r)}\left\{\mathsf{x}\right\} - r_q = \int_{-\infty}^{\infty} x\, p_{\mathsf{x}}\left(x|\mathsf{r} = r_q - y\right)\, \mathrm{d}x - r_q. \tag{50}$$

Thus, to solve the estimation problem, the conditional PDF $p_{\mathsf{x}}\left(x|\mathsf{r} = r_q - y\right)$ must be known. It is assumed that $p_{\mathsf{x}}\left(x|\mathsf{r} = r_q - y\right)$ is independent from r_q, which is approximately valid for AWGN attacks and an almost flat PDF $p_{\mathsf{x}}(x)$ in the range of one quantization interval, e.g., fine quantization, so that $\hat{q}(r_q, y) = \hat{q}(y)$. Thus, the random variable y with support in $[-\frac{\Delta}{2}, \frac{\Delta}{2})$ is introduced and the PDF $p_{\mathsf{x}}\left(x|\mathsf{r} = r_q - y\right) = p_{\mathsf{x}}\left(x|\mathsf{y} = y\right)$ is considered in the following.

First, Bayes' rule is applied which yields

$$p_{\mathsf{x}}\left(x|\mathsf{y} = y\right) = \frac{p_{\mathsf{x}}(x)\, p_{\mathsf{y}}\left(\mathsf{y} = y|\mathsf{x} = x\right)}{p_{\mathsf{y}}\left(\mathsf{y} = y\right)}, \quad \text{for} \quad -\infty < x < \infty. \tag{51}$$

Next, $p_{\mathsf{y}}\left(\mathsf{y} = y|\mathsf{x} = x\right)$ has to be computed. Due to the quantization involved in the embedding procedure, an elegant closed-form of $p_{\mathsf{y}}\left(\mathsf{y} = y|\mathsf{x} = x\right)$ does not exist. Another difficulty is the unlimited support of the random variable x. However, it turns out that, for $\sigma_{\mathsf{v}}^2 \leq \sigma_{\mathsf{v},\text{design}}^2$ and $r_q = 0$, a sufficiently accurate approximation is obtained by considering only $x \in [-\frac{3\Delta}{2}, \frac{3\Delta}{2})$.

With this approximation, the assumption of white Gaussian attack noise v of power σ_{v}^2 and the numerical represenation of $p_{\mathsf{y}}(y|d)$, (51) can be evaluated, leading to a numerical represenation of $p_{\mathsf{x}}\left(x|\mathsf{y} = y\right)$. Applying these result to (50) yields the desired estimate $\hat{q}(y)$. For illustration purpose, Fig. 21 depicts the PDFs of y and the resulting $\hat{q}$ for $\text{WNR} = \text{WNR}_{\text{design}} = 0$ dB.

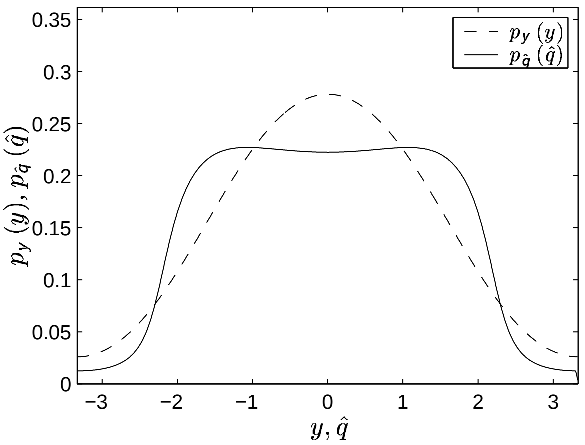


Fig. 21. PDFs of received extracted data before and after inverse SCS mapping ($\mathrm{WNR} = 0$ dB)

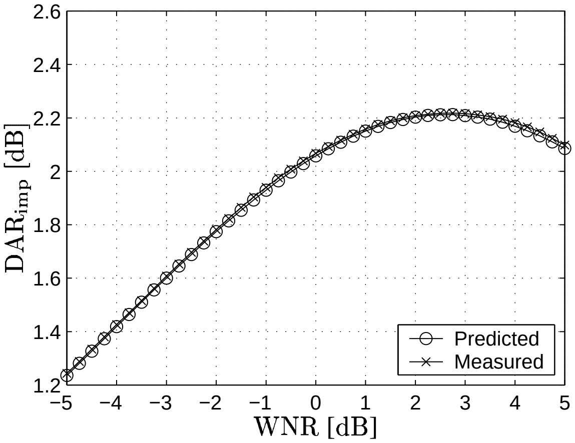


Fig. 22. Distortion improvement for inverse SCS mapping after AWGN attack with $\sigma_v^2 = \sigma_{v,\mathrm{design}}^2/4$.

B.2 Achievable Distortion Reduction

Finally, the achieved distortion improvement is investigated. The improvement is measured in terms of the difference between the DAR (document-to-attack power ratio) before (DAR_r) and after ($\mathrm{DAR}_{\hat{x}}$) the mapping, which is given by

$$\mathrm{DAR_{imp}} = \mathrm{DAR}_{\hat{x}} - \mathrm{DAR}_r = 10\log_{10}\frac{\mathrm{D}\left(\mathbf{x},\mathbf{r}\right)}{\mathrm{D}\left(\mathbf{x},\hat{\mathbf{x}}\right)}\ \mathrm{dB}. \tag{52}$$

Unfortunately, the result for $\mathrm{WNR} = \mathrm{WNR_{design}}$ is rather disappointing with a maximal distortion improvement of $\mathrm{DAR_{imp,max}} = 0.04$ dB. This value has been obtained via simulations and numerical evaluation of $\mathrm{E}_{p_x(x|r)}\left\{(\hat{x}-x)^2\right\}$. Obviously, the optimal quantizer step size in SCS is such that, after AWGN attacks, the watermark embedding distortion is no longer invertible. Fig. 22 depicts the same result for the case of the SCS quantizer step size design for a noise power being 6 dB above the given channel noise power. In this case, the maximum distortion improvement is about 2.2 dB. Although this improvement might be of interest in practice, it is important to emphasize that such an improvement could be obtained only for very mild channel conditions. Note that the dependency of $\mathrm{DAR_{imp}}$ on the WNR is due to the variable choice of Δ depending on the WNR.

It has to be concluded that the inversion of SCS watermarking after AWGN attacks is practically impossible or at least inefficient. Nevertheless, the derived inverse SCS mapping might be useful. Suppose the owner of a signal stores only the SCS watermarked version and erases the original. In this case, the SCS watermark might be designed for strong robustness, that is, low WNRs. However, even without an explicit attack, the watermarked signal is slightly distorted due to quantization, which might occur when storing the data. This quantization can be approximated by low-power noise. In such a scenario, the inverse scaling derived for the noiseless case might be not appropriate, but the MMSE estimation removes a good deal of the distortion introduced by the SCS watermark, as demonstrated in Fig. 22.

VIII. Conclusions and Outlook

Information embedding into IID original data and an attack by AWGN has been investigated. The decoder has no access to the original data. This scenario can be considered communication with side-information at the encoder for that a theoretical communication scheme has been derived by Costa in 1983. In this paper, a suboptimal practical version of Costa's scheme has been studied. The new scheme is named "scalar Costa scheme" (SCS) due to the involved scalar quantization during encoding and decoding. A performance comparison of different blind watermarking schemes shows that SCS outperforms the related DM techniques for low WNRs and performs significantly better than state-of-the-art blind SS watermarking for the relevant range of WNRs. The latter result is mainly due to the independence of SCS from the characteristics of the original signal. SCS combined with coded modulation achieves a rate of 1 bit/element at $\mathrm{WNR} \geq 8.3$ dB, which is within 1.6 dB of the SCS capacity. For $\mathrm{WNR} \approx 1.5$ dB, SCS communication with rate 1/3 turbo coding achieves $\mathrm{BER} < 10^{-5}$. For lower WNRs, SCS should be combined with the spread-transform (ST) technique so that SCS operates effectively at a $\mathrm{WNR} \approx 1.5$ dB.

Two further topics that are relevant for the usage of SCS in practical information hiding systems are investigated. These are the robustness to amplitude scaling on the watermark channel and the removal of watermark embedding distortion by authorized parties. Robustness against amplitude scaling can be achieved via robust estimation of the proper SCS quantizer step size at the receiver as described in Sec. VI. In Sec. VII, it is shown that the reduction of watermark embedding distortion is possible for low attack noise.

The performance gap between SCS and ICS has to be bridged by constructing more complicated codebooks and by extending the embedding and detection rule to non-scalar operations. Research in this direction has been started, e.g., by Chou et al. [18]. However, SCS might still remain an attractive technique for many information embedding applications due to its simple structure and host signal independent design.

Acknowledgments

The authors would like to thank Jonathan Su for the contributions to this work during his stay at the University of Erlangen-Nuremberg. Further, we would like to thank Ton Kalker for fruitfull discussions concerning the SCS step size estimation which helped to improve the explanation of our alorithm significantly.

References

[1] B. Chen and G. W. Wornell, "Provably robust digital watermarking," in *Proceedings of SPIE: Multimedia Systems and Applications II (part of Photonics East '99)*, Boston, MA, USA, September 1999, vol. 3845, pp. 43–54.

[2] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Transaction on Information Theory*, January 2001.

[3] J. K. Su and B. Girod, "Power-spectrum condition for energy-efficient watermarking," in *Proceedings of the IEEE Intl. Conference on Image Processing 1999 (ICIP '99)*, Kobe, Japan, October 1999.

[4] J. J. Eggers, J. K. Su, and B. Girod, "Performance of a practical blind watermarking scheme," in *Proc. of SPIE Vol. 4314: Security and Watermarking of Multimedia Contents III*, San Jose, Ca, USA, January 2001.

[5] Joachim J. Eggers, *Information Embedding and Digital Watermarking as Communication with Side Information*, Ph.D. thesis, Lehrstuhl für Nachrichtentechnik I, Universität Erlangen-Nürnberg, Erlangen, Germany, November 2001.

[6] J. Eggers and B. Girod, *Informed Watermarking*, Kluwer Academic Publishers, Boston, Dordrecht, London, 2002.

[7] B. Chen and G. W. Wornell, "Achievable performance of digital watermarking systems," in *Proceedings of the IEEE Intl. Conference on Multimedia Computing and Systems (ICMCS '99)*, Florence, Italy, June 1999, vol. 1, pp. 13–18.

[8] I. J. Cox, M. L. Miller, and A. L. McKellips, "Watermarking as communications with side information," *Proceedings of the IEEE, Special Issue on Identification and Protection of Multimedia Information*, vol. 87, no. 7, pp. 1127–1141, July 1999.

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

[10] B. Chen and G. W. Wornell, "Digital watermarking and information embedding using dither modulation," in *Proc. of IEEE Workshop on Multimedia Signal Processing (MMSP-98)*, Redondo Beach, CA, USA, Dec. 1998, pp. 273–278.

[11] B. Chen and G. Wornell, "Preprocessed and postprocessed quantization index modulation methods for digital watermarking," in *Proc. of SPIE Vol. 3971: Security and Watermarking of Multimedia Contents II*, San Jose, Ca, USA, January 2000, pp. 48–59.

[12] M. Ramkumar and A. N. Akansu, "Self-noise suppression schemes in blind image steganography," in *Proceedings of SPIE: Multimedia Systems and Applications II (part of Photonics East '99)*, Boston, MA, USA, September 1999, vol. 3845, pp. 55–65.

[13] M. Ramkumar, *Data Hiding in Multimedia: Theory and Applications*, Ph.D. thesis, Dep. of Electrical and Computer Engineering, New Jersey Institute of Technology, Kearny, NJ, USA, November 1999.

[14] M. Ramkumar and A. N. Akansu, "FFT based signaling for multimedia steganography," in *Proceedings of the IEEE Intl. Conference on Speech and Signal Processing 2000 (ICASSP 2000)*, Istanbul, Turkey, June 2000.

[15] M. Ramkumar and A. N. Akansu, "Floating signal constellations for multimedia steganography," in *Proceedings of the IEEE International Conference on Communications, ICC 2000, New Orleans, LA, USA*, June 2000, vol. 1, pp. 249–253.

[16] L. Gang, A. N. Akansu, and M. Ramkumar, "Periodic signaling scheme in oblivious data hiding," in *Proceedings of 34th Asilomar Conf. on Signals, Systems, and Computers*, Asilomar, CA, USA, October 2000.

[17] J. J. Eggers, J. K. Su, and B. Girod, "A blind watermarking scheme based on structured codebooks," in *Secure Images and Image Authentication, Proc. IEE Colloquium*, London, UK, April 2000, pp. 4/1–4/6.

[18] J. Chou, S. Pradhan, L. El Ghaoui, and Kannan Ramchandran, "A robust optimization solution to the data hiding problem using distributed source coding principles," in *Proc. of SPIE Vol. 3974: Image and Video Communications and Processing 2000*, San Jose, Ca, USA, January 2000.

[19] J. Chou, K. Ramchandran, and S. Pradhan, "Turbo coded trellis-based constructions for data hiding," in *Proc. of SPIE: Security and Watermarking of Multimedia Contents IV*, San Jose, Ca, USA, January 2002.

[20] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice-strategies for cancelling known interference," in *ISITA 2000, Honolulu, HI, USA*, November 2000, pp. 681–684.

[21] S. Shamai and B. M. Zaidel, "Applying the dirty paper approach to the downlink of a cellular system," Tech. Rep., Technion-Israel Institute of Technology, 2002, In Preparation.

[22] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1250–1276, June 2002.

[23] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, Springer-Verlag, New York, Berlin, Heidelberg, 1988.

[24] L. Schuchman, "Dither signals and their effect on quantization noise," *IEEE Transaction on Communication Technology (COM)*, vol. 12, pp. 162–165, December 1964.

[25] R. M. Gray and T. G. Stockham, "Dithered quantizers," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 805–812, May 1993.

[26] B. Chen and G. W. Wornell, "Dither modulation: a new approach to digital watermarking and information embedding," in *Proc. of SPIE Vol. 3657: Security and Watermarking of Multimedia Contents*, San Jose, January 1999, pp. 342–353.

[27] B. Chen and G. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. on Information Theory*, vol. 47, no. 4, May 2001.

[28] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[29] C. Heegard and A.A. El Gamal, "On the capacity of computer memory with defects," *IEEE Transactions on Information Theory*, vol. 29, no. 5, pp. 731–739, September 1983.

[30] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, 1991.

[31] F. Hartung and B. Girod, "Digital watermarking of raw and compressed video," in *Proceedings EUROPTO/SPIE European Conference on Advanced Imaging and Network Technologies*, Berlin, Germany, October 1996.

[32] I. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673–1687, 1997.

[33] B. Chen, *Design and Analysis of Digital Watermarking, Information Embedding, and Data Hiding Systems*, Ph.D. thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Boston, MA, USA, June 2000.

[34] J. Chou, K. Ramchandran, and A. Ortega, "Next generation techniques for robust and imperceptible audio data hiding," in *Proceedings of the IEEE Intl. Conference on Speech and Signal Processing 2001 (ICASSP 2001)*, Salt Lake City, Utah, USA, May 2001.

[35] G. Ungerboeck, “Channel coding with multilevel/phase signals,” *IEEE Transactions on Information Theory*, vol. 28, no. 1, pp. 55–67, January 1982.

[36] C. Berrou and A. Glavieux, “Near optimum error correcting coding and decoding,” *IEEE Transactions on Communications*, vol. 44, no. 10, pp. 284–287, October 1996.

[37] J. J. Eggers, R. Bäuml, and B. Girod, “Estimation of amplitude modifications before SCS watermark detection,” in *Proc. of SPIE: Security and Watermarking of Multimedia Contents IV*, San Jose, Ca, USA, January 2002.

[38] B. Macq and F. Dewey, “Trusted headers for medical images,” in V^3D^2 *Watermarking Workshop*, Erlangen, Germany, October 1999.