Quantization Effects on Digital Watermarks

Joachim J. Eggers [a], Bernd Girod [b]

[a] *Telecommunications Laboratory, University of Erlangen-Nuremberg*
Cauerstr. 7/NT, 91058 Erlangen, Germany, eggers@LNT.de

[b] *Information Systems Laboratory, Stanford University*
Stanford, CA 94305-9510, USA, girod@ee.stanford.edu

Abstract

Digital watermarking is the process of introducing small modifications into a copy of a digital document that can be detected later. The embedded information can be used to determine the document's owner or simply to distinguish several copies. However, coincidental or malicious "attacks" can degrade the robustness of watermark detection. Here, uniform scalar quantization of watermarked documents is investigated theoretically, extending results from theory of dithered quantization, and experimentally. The watermark is embedded by an independent additive pseudo-noise sequence. The statistical distribution of the quantization errors depending on the statistics of the host signal and the watermark is used to determine the robustness of watermark detection via correlation. Experiments with JPEG compression of an image with a DCT-domain additive watermark demonstrate the usefulness of the presented theory.

Mit Einbettung von digitalen Wasserzeichen bezeichnet man den Prozeß zum Einbringen kleiner Veränderungen in eine Kopie eines digitalen Dokumentes, welche später detektiert werden können. Die eingebettete Information kann benutzt werden, um den Besitzer des Dokumentes zu bestimmen oder einfach um verschiedene Kopien zu unterscheiden. Allerdings können zufällige oder böswillige Angriffe die Robustheit der Wasserzeichendetektion reduzieren. Hier wird die gleichmäßige skalare Quantisierung von mit Wasserzeichen versehenen Dokumenten theoretisch und experimentell untersucht. Die theoretische Untersuchung basiert auf der Erweiterung der Theorie zur geditherten Quantisierung. Das Wasserzeichen wird mittels einer unabhängigen additiven Pseudo-Rauschsequenz eingebettet. Die statistische Verteilung des Quantisierungsfehlers in Abhängigkeit von den statistischen Eigenschaften des Originalsignals und des Wasserzeichens wird zur Bestimmung der Robustheit der Wasserzeichendetektion mittels Korrelation herangezogen. Experimente mit JPEG Kompression eines Bildes mit im DCT-Bereich eingebetteten additiven Wasserzeichen demonstrieren die Nützlichkeit der vorgestellten Theorie.

Key words: Digital watermarking; Image watermarking; Watermark robustness; Quantization; Dithered quantization; JPEG-compression

1 Introduction

The digital representation of audio, image and video signals has become very popular in the last decade. The success of the digital technology is mainly due to the possibility of efficient transmission and copying. However, unauthorized copying is also simplified. One approach to combat this problem is to mark a digital document such that a copyright can be proven or the distribution path be traced. The marking process produces a perceptually equivalent digital document rather than a bit-exact copy.

For a general watermarking scheme, the embedding process can be described by

$$\vec{s}_k = \vec{x} + \vec{w}_k, \tag{1}$$

where $\vec{x}$ denotes the original signal, $\vec{w}_k$ the signal modification introduced by the watermarking process, and $\vec{s}_k$ the published signal (watermarked document). $\vec{x}$ is also called "host signal" or "private document". In the remainder of this article, signals are denoted by vectors (e.g., $\vec{x}$), the nth signal sample by $x[n]$, and random variables by boldface (e.g., $\mathbf{x}$). The index k is used to distinguish between different watermarks.

Note that (1) can describe a general watermarking scheme since we define the watermark to be the difference between the host signal $\vec{x}$ and the published signal $\vec{s}$, which is possible in all watermarking schemes where the signal size is not changed in the embedding process. This does not necessarily mean that (1) is a very useful definition in all cases, e.g., for embedding schemes based on geometrical deformations. Here, we consider watermark embedding by the simple addition of a pseudo-noise sequence $\vec{w}_k$ which is statistically independent from the host signal $\vec{x}$ and from other possible watermarks. That is, $\vec{w}_i$ and $\vec{w}_j$, with $j \neq i$, are independent of each other. In this case, other watermarks than the one to be detected appear as additive noise. Thus, the watermark index k is neglected in the remainder. It is well known that the watermark power should be adapted to the host signal to ensure imperceptibility of the watermark. We propose in [1] to separate the host signal into sub-signals such that the allowed watermark power and maximal strength of the attack is constant within one sub-signal. From the point of view of watermark transmission each sub-signal forms one sub-channel. If the detection performance for each sub-channel is known, the outputs of all sub-channels can be combined for maximally robust detection. In this paper we focus on analyzing the detection performance for one sub-channel, thus, a stationary watermark $\vec{w}$ is assumed. The concept of sub-channels is considered again in Section 5.

The watermark detector in general receives a signal

$$\vec{r} = \vec{s} + \vec{e} = \vec{x} + \vec{w} + \vec{e}, \tag{2}$$

where $\vec{e}$ can include any distortion that might be introduced by the watermark channel. (2) can describe a general attack, however, as with (1) this formulation is not very useful in some cases, e.g., desynchronization attacks. (2) is applicable in case of quantization or denoising attacks, but statistical dependencies between $\vec{e}$, $\vec{w}$ and $\vec{x}$ must be considered.

Depending on the system architecture encapsulating the watermarking scheme the host signal may or may not be available to the watermark detector. There are schemes where the private signal is provided by some trusted third party, but in other cases, e.g., in copy-protection for Digital Versatile Discs (DVD), watermark detection must be possible without reference to the private document $\vec{x}$. Costa [2] and Moulin and O'Sullivan [3] show that for Gaussian signals under certain linearity assumptions that allow an information-theoretic analysis watermarking schemes that do not provide the host signal $\vec{x}$ at the detector can perform exactly as well as schemes where $\vec{x}$ is not available at the detector. Chen and Wornell [4] point out that for AWGN attacks the watermark capacity is independent of the host signal statistics. However, when the detector has no access to the host signal, the embedding process must be adapted appropriately. Practical schemes based on this philosophy are described in [4–7]. In this paper, we consider only watermark embedding by an independent additive sequence $\vec{w}$, thus, the discussed scheme cannot be optimal when the host signal is not available to the watermark detector. In the following discussion, the interference from the private signal $\vec{x}$ is reduced by subtracting $\vec{x}$, weighted by a factor γ_x, from the received signal $\vec{r}$. The detection process is performed using the pre-processed signal

$$\vec{y} = \vec{r} - \gamma_x \vec{x}. \tag{3}$$

When $\vec{x}$ is available to the watermark detector, γ_x is chosen such that the correlation between the host signal $\vec{x}$ and the pre-processed received signal $\vec{y}$ is completely removed. In all other cases, host signal interference reduction methods, e.g., host signal estimation, should be implemented. We do not specify such methods in this paper since several additional concepts are necessary to describe these schemes. However, the effect of host signal reduction is roughly reflected in our analysis by the weighted subtraction of $\vec{x}$, where γ_x will be smaller for less effective reduction methods and $\gamma_x = 0$ when no reduction is implemented. We denote the detection case with $\gamma_x = 0$ as “blind detection”.

The complete characterization of the watermark channel – equivalent to the description of $\vec{e}$ – is still an open problem. In contrast to many other communications problems, the channel distortion $\vec{e}$ might be introduced intentionally to remove or hide the transmitted information; the watermark is attacked. The range of possible attacks is constrained by the maximum allowed document distortion. The watermark need not be detectable from a worthless document. However, this bound is difficult to describe exactly due to the subjective quality rating for natural data. In addition, this distortion bound is also important for the watermark embedding, since the watermark must not degrade the document quality either. Correct modeling of the worst-case channel is crucial for the proper design of watermarks.

The most common approach for verifying watermark robustness is to investigate *experimentally* the detection performance in the presence of a certain watermark channel. In [8–12], image watermark detection after JPEG compression is investigated. Blurring, image rotation, scaling and cropping are other frequently used attacks [8,10,13]. Kutter and Petitcolas [11] present a benchmark test for image watermarking schemes based on performance tests for many different attacks. Experimental investigations of image watermark robustness should follow the proposed scheme to make results comparable.

Experimental verifications allow a comparison between different schemes, where even subjective quality ratings can be considered. However, these experiments give little information about possible improvements of an investigated scheme, and statements about the maximum robustness can hardly be made. Therefore, a theoretical investigation of the watermark channel is important.

Theoretical investigations of watermark detection are given in [14,9,15–17]. Most of the presented results are valid when the channel distortion can be approximated by independent additive white Gaussian noise (AWGN). We will analyze the effects of uniform scalar quantization, as it is often used in lossy compression schemes. *In general, the distortion introduced by quantization cannot be modeled appropriately by AWGN.* In the case of correlation detection, which is investigated in this paper, a significant reduction of watermark correlation has to be considered. This will be shown in Section 4.

Our analysis of watermark detection after quantization attacks is based on a correlation detector, where the decision boundary can be adapted to the channel characteristics. This detector is described in Section 2. In order to describe the robustness of watermark detection after quantization, we need an analytical expression for the dependencies between the watermarked signal $\vec{s}$ and the quantized watermarked signal $\vec{r}$. This can be found when considering additive watermark embedding followed by quantization as *dithered quantization*. In Section 3 the fundamentals of dithered quantization are briefly reviewed, and the theory of dithered quantizers is extended as necessary for our purpose. In Section 4 the results are applied to investigate theoretically watermark detection using correlation measurements after uniform scalar quantization. The detection performance for host signals and watermark signals with different distributions is compared, where the quantizer step size Δ is varied. In Section 5, an example image watermarking scheme is presented. The detection performance for watermark components after JPEG compression is measured experimentally and also predicted using the theory described in Section 4. Section 6 concludes the presented investigation of watermark robustness in the presence of quantization attacks.

2 Watermark Detection

Signal detection has been analyzed extensively by communication engineers. Here, we only summarize the aspects that are important for watermark detection. We assume that the watermark detector is always synchronized. At the first glance this assumption seems very restrictive. However, as shown in [18,19], it is possible to counterattack many desynchronization attacks when using an improved detector or applying some re-synchronization prior to the watermark detection.

2.1 Bayes' Hypothesis Test

The watermark detection problem can be stated as a simple-hypothesis test, with the hypotheses

$$\begin{aligned} \mathrm{H}_0 &: \text{ the watermark } \vec{w} \text{ is not present,} \\ \mathrm{H}_1 &: \text{ the watermark } \vec{w} \text{ is present.} \end{aligned}$$

The task of hypothesis testing is to decide for a pre-processed received document $\vec{y}$, which of the hypotheses is true. Usually, it is not possible to separate all watermarked and un-watermarked documents perfectly. We have to trade off the probability p_{FP} of accepting H_1 when H_0 is true (*false positive*) and the probability p_{FN} of accepting H_0 when H_1 is true (*false negative*). Bayes' solution is the decision rule

$$\mathrm{H}_1 : \ \frac{p_{\mathbf{y}}\left(\vec{y}|\mathrm{H}_1\right)}{p_{\mathbf{y}}\left(\vec{y}|\mathrm{H}_0\right)} > K = \frac{\text{(cost false positive)} \ \cdot p_{\mathrm{H}_0}}{\text{(cost false negative)} \ \cdot p_{\mathrm{H}_1}}, \tag{4}$$

where K is a constant depending on the a priori probabilities for H_1 and H_0 and the cost connected with the different decision errors [20]. The costs for false positive and false negative errors are introduced to weight the severity of false positive and false negative errors depending on the actual application. For $K = 1$, the decision rule (4) forms a **maximum-likelihood (ML) detector**. For equal a priori probabilities, the overall detection error probability is $p_e = \frac{1}{2}(p_{\mathrm{FP}} + p_{\mathrm{FN}})$. Receiver operating characteristic (ROC) graphs, as proposed in [11], can be computed using different thresholds K.

The probability density functions (PDFs) $p_{\mathbf{y}}\left(\vec{y}|\mathrm{H}_1\right)$ and $p_{\mathbf{y}}\left(\vec{y}|\mathrm{H}_0\right)$ must be known for the implementation of (4). Therefore, statistical models are used to design the optimal decision rule and to estimate the corresponding error probabilities. Two such models are described in the following sub-sections.

2.2 AWGN Channel

A simple approach is to model the channel distortion by additive white Gaussian noise (AWGN). For watermark applications, this implies that the original signal $\mathbf{x}$ and the channel distortion $\mathbf{e}$ are jointly Gaussian random processes and statistically independent from a possibly included watermark. The AWGN channel can be analyzed completely by the detection performance for one watermark realization $\vec{w}$. Here, $p_{\mathbf{y}}\left(\vec{y}|\mathrm{H}_1\right)$ and $p_{\mathbf{y}}\left(\vec{y}|\mathrm{H}_0\right)$ are Gaussian PDFs with equal variance but different means, and the test (4) becomes a common correlation detector [21]:

$$\mathrm{H}_1 : \quad C = \frac{(\vec{r} - \gamma_x \vec{x})^T \vec{w}}{||\vec{w}||^2} = \frac{\vec{y}^T \vec{w}}{||\vec{w}||^2} > \frac{1}{2} + \frac{\sigma_o^2 \log K}{||\vec{w}||^2}, \tag{5}$$

where σ_o^2 denotes the variance of the AWGN. The corresponding decision errors for $K = 1$ are depicted in Fig. 1 (a). The solution is similar for non-white signals when pre-filtering is applied.

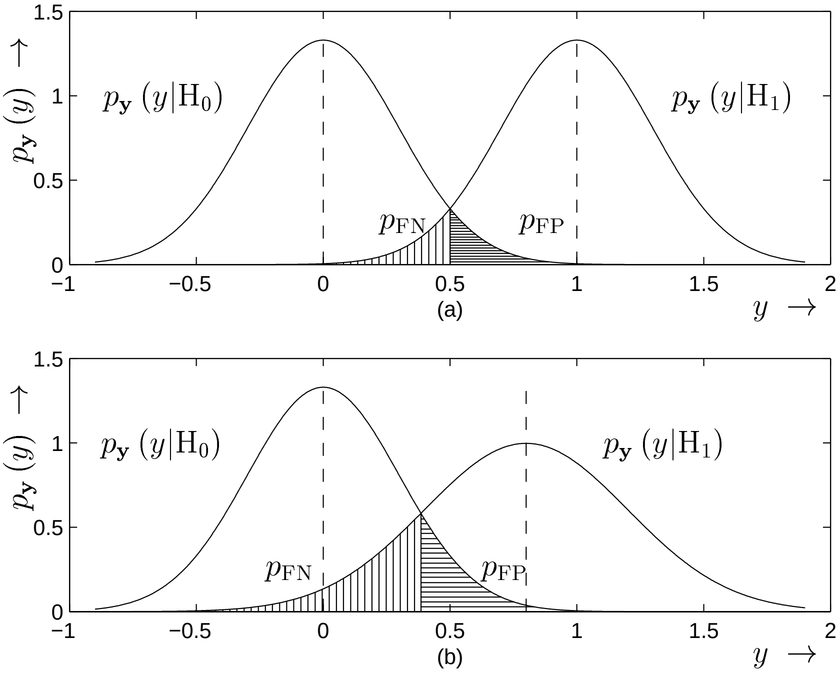

Fig. 1. (a) Correlation detection for an AWGN channel. $\mathrm{E}\left\{\mathbf{y}|\mathrm{H}_0\right\}$=0; $\mathrm{E}\left\{\mathbf{y}|\mathrm{H}_1\right\}$=1; $\sigma_o^2 = \mathrm{Var}\left\{\mathbf{y}|\mathrm{H}_0\right\} = \mathrm{Var}\left\{\mathbf{y}|\mathrm{H}_1\right\} = 0.3^2$; $K = 1$. (b) Correlation detection for signal dependent channel distortion. $\mathrm{E}\left\{\mathbf{y}|\mathrm{H}_0\right\}$=0; $\mathrm{E}\left\{\mathbf{y}|\mathrm{H}_1\right\}$=0.8; $\mathrm{Var}\left\{\mathbf{y}|\mathrm{H}_0\right\} = 0.3^2$; $\mathrm{Var}\left\{\mathbf{y}|\mathrm{H}_1\right\} = 0.4^2$; $K = 1$.

The quantization channel, as discussed in Section 3 and Section 4, introduces distortions that depend in general on the quantizer input. Due to this dependency, *this channel can only be described by its average behavior for all possible watermarks,* so the watermark should be viewed as a random variable, too. The pre-processed received signal is dependent on the given hypothesis:

$$\mathrm{H}_0: \quad \vec{\mathbf{y}} = (1-\gamma_x)\vec{\mathbf{x}} + \vec{\mathbf{e}}_{\mathrm{H}_0} \tag{6}$$
$$\mathrm{H}_1: \quad \vec{\mathbf{y}} = (1-\gamma_x)\vec{\mathbf{x}} + \vec{\mathbf{e}}_{\mathrm{H}_1} + \vec{\mathbf{w}}. \tag{7}$$

Here, a correlation detector need not be optimal in general. The derivation of the correlation detector starting from a hypothesis test, as shown in [21], is only valid for independent AWGN. Nevertheless, the correlation between the pre-processed received signal $\vec{y}$ and the watermark $\vec{w}$ can be used as a similarity measurement. Correlation detection is known to be a good candidate when accurate statistical models for the channel distortion are not available [17]. To describe the correlation detector, we define the sample-wise product

$$c[n] = y[n]w[n]/\sigma_{\mathbf{w}}^2, \tag{8}$$

where $\sigma_{\mathbf{w}}^2$ denotes the variance of the watermark. In the remainder of this article, we assume $\vec{c}$ to be ergodic and model the elements $c[n]$ by a random variable $\mathbf{c}$. The expectation $\mathrm{E}\{\mathbf{c}\}$ is equal to the normalized correlation of the watermark with the pre-processed received signal ($\mathrm{E}\{\mathbf{c}\} = \mathrm{E}\left\{\vec{\mathbf{y}}^T\vec{\mathbf{w}}\right\}/M\sigma_{\mathbf{w}}^2$), and thus indicates whether the watermark is embedded or not. For finite length signals, we estimate this expectation via

$$\mathrm{E}\{\mathbf{c}\} \approx C = \frac{1}{M}\sum_{n=1}^{M} c[n], \tag{9}$$

where M is the number of samples considered for the estimate. $\mathbf{y}$ and $\mathbf{w}$ may be non-Gaussian, but they describe IID random processes. Thus, we can apply the central limit theorem, and for sufficiently large M, this estimate can be modeled by a Gaussian random variable $\mathbf{C}$ with variance [22]

$$\mathrm{Var}\{\mathbf{C}\} = \frac{1}{M}\mathrm{Var}\{\mathbf{c}\}. \tag{10}$$

Assuming $\mathbf{C}$ to be Gaussian, we can implement (4). We expect $\mathrm{E}\{\mathbf{c}|\mathrm{H}_0\} = 0$. The values $\mathrm{E}\{\mathbf{c}|\mathrm{H}_1\}$, $\mathrm{Var}\{\mathbf{c}|\mathrm{H}_1\}$ and $\mathrm{Var}\{\mathbf{c}|\mathrm{H}_0\}$ are dependent on the watermark channel. Based on these parameters, the optimal decision rule is

$$\mathrm{H}_1: \quad K < \frac{(2\pi \text{ Var}\{\mathbf{c}|\mathrm{H}_1\}/M)^{-\frac{1}{2}} \exp\left(-\frac{(C-\mathrm{E}\{\mathbf{c}|\mathrm{H}_1\})^2}{2 \text{ Var}\{\mathbf{c}|\mathrm{H}_1\}/M}\right)}{(2\pi \text{ Var}\{\mathbf{c}|\mathrm{H}_0\}/M)^{-\frac{1}{2}} \exp\left(-\frac{(C-0)^2}{2 \text{ Var}\{\mathbf{c}|\mathrm{H}_0\}/M}\right)} \tag{11}$$

or equivalently

$$\begin{aligned} \mathrm{H}_1: \; & 2\frac{\mathrm{E}\{\mathbf{c}|\mathrm{H}_1\} C}{\text{Var}\{\mathbf{c}|\mathrm{H}_1\}} + \left(\frac{1}{\text{Var}\{\mathbf{c}|\mathrm{H}_0\}} - \frac{1}{\text{Var}\{\mathbf{c}|\mathrm{H}_1\}}\right) C^2 \\ & > \frac{1}{M}\left(2\log(K) - \log\frac{\text{Var}\{\mathbf{c}|\mathrm{H}_0\}}{\text{Var}\{\mathbf{c}|\mathrm{H}_1\}}\right) + \frac{\mathrm{E}\{\mathbf{c}|\mathrm{H}_1\}^2}{\text{Var}\{\mathbf{c}|\mathrm{H}_1\}}. \end{aligned} \tag{12}$$

In Fig. 1 (b), the corresponding detection situation is depicted for $K = 1$. Besides the unequal variances Var $\{\mathbf{c}|\mathrm{H}_1\}$ and Var $\{\mathbf{c}|\mathrm{H}_0\}$, the decrease of $\mathrm{E}\{\mathbf{c}|\mathrm{H}_1\}$ compared to the AWGN case is shown. For equal variances Var $\{\mathbf{c}\}$ = Var $\{\mathbf{c}|\mathrm{H}_1\}$ = Var $\{\mathbf{c}|\mathrm{H}_0\}$, the quadratic term in (12) vanishes and the rule simplifies to

$$\mathrm{H}_1: \; C > \frac{\mathrm{E}\{\mathbf{c}|\mathrm{H}_1\}}{2} + \frac{\text{Var}\{\mathbf{c}\}\log(K)}{M\mathrm{E}\{\mathbf{c}|\mathrm{H}_1\}}, \tag{13}$$

which is very similar to (5). The important difference is that (13) is dependent on $\mathrm{E}\{\mathbf{c}|\mathrm{H}_1\}$. When using (13) with $K = 1$, the detection error probabilities p_{FP} and p_{FN} are equal and can be computed with

$$p_{\mathrm{FP}} = p_{\mathrm{FN}} = \frac{1}{2}\text{erfc}\left(\sqrt{M}\frac{\mathrm{E}\{\mathbf{c}|\mathrm{H}_1\}}{2\sqrt{2\text{Var}\{\mathbf{c}\}}}\right) \tag{14}$$

where erfc $(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-\xi^2) \; \mathrm{d}\xi$. Due to the Gaussian model for $\mathbf{C}$, p_{FP} and p_{FN} can be computed similarly for detection schemes using (12) and for $K \neq 1$. Note that the performance of the described detection method can be characterized completely by $\mathrm{E}\{\mathbf{c}|\mathrm{H}_1\}$, Var $\{\mathbf{c}|\mathrm{H}_1\}$ and Var $\{\mathbf{c}|\mathrm{H}_0\}$.

2.4 Improved Detection for Known Signal and Channel Statistics

In the previous subsection the correlation detection is motivated as a robust detection method when little is known about the statistics of the detection interference. Particular when detecting without knowing the host signal $\vec{x}$, knowledge about the statistics of the host signal can be exploited for improved blind detection. In Section 5, we introduce the generalized Gaussian model which can describe the statistics of DCT coefficients of images very accurately. Hernandez [17] derives a blind detection rule for host signals with a generalized Gaussian PDF which differs significantly from a correlation measurement. Such a detection will be very useful for blind detection after weak quantization attacks where the host signal interference dominates the quantization noise. However, for coarse quantization the statistical model must be adapted, which is not straightforward. In addition the analysis of quantization effects presented in this paper is not sufficient to describe theoretically the performance of such detection after quantization attacks. For this reason only correlation detection is considered in the remainder. Extending the robustness analysis to improved detectors, e.g., the one described in [17], might be a topic for future research. However, most probably it is more useful to analyze schemes where the host signal interference is already suppressed by a proper embedding process [3–7]. The results for blind detection, presented in this paper, can serve as a lower bound on the achievable detection performance. For detection with original the correlation detection is a good choice anyway, since an accurate modeling of all possible attacks might be difficult in general.

3 Dithered Quantization

In Section 4, we will show that quantization of a watermarked document can be considered as dithered quantization of the original signal. This point of view enables the theoretical analysis of the robustness of watermark correlation detection after quantization attacks. Due to the central importance of dithered quantizers for our analysis, we briefly review the fundamentals of dithered quantization before extending the theory to aspects important for watermarking schemes. Previous work on dithered uniform scalar quantizers can be found, for instance, in [23–25].

3.1 Fundamentals of Dithered Quantization

A dithered quantizer is a quantizer that adds a dither sequence $\vec{d}$ to the input signal $\vec{x}$ before discretizing the samples. There are two kinds of dithered quantizers, the subtractive dithered quantizer as depicted in Fig. 2, and the non-subtractive dithered quantizer. Non-subtractive dithered quantizers become important, when the receiver of the digital data has no synchronized access to the dither sequence. However, the distortion introduced by a subtractive dithered quantizer is lower, which can be seen easily from the expressions for the quantization error:

$$\text{subtractive dithered quantizer} \; : \; \vec{e} = \underline{\underline{\vec{z}}} - \vec{x} = \underline{\vec{z}} - \vec{z} \tag{15}$$

$$\text{non-subtractive dithered quantizer} \; : \; \vec{\epsilon} = \underline{\vec{z}} - \vec{x} = \vec{e} + \vec{d}, \tag{16}$$

where $\vec{\underline{z}}$ denotes the output of the non-subtractive dithered quantizer, and $\vec{\underline{\underline{z}}}$ denotes the output of the subtractive dithered quantizer.

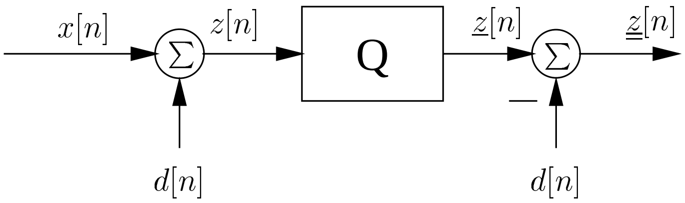

Fig. 2. Subtractive dithered quantizer

3.1.1 Previous Results

The investigation of dithered quantizers was mainly motivated by the goal of obtaining quantization errors independent from the original signal. This is important for extending the analysis of fine quantization to coarse quantization, independent from the signal to be quantized, or to achieve improved subjective quantization quality. In watermarking schemes that embed additive signals $\vec{w}$, the watermark $\vec{w}$ can be considered as a dither signal $\vec{d}$. Thus, we are interested in the correlation between the quantized signal $\vec{\underline{\underline{z}}}$ and the dither signal $\vec{d}$ dependent on the characteristics of the input signal $\vec{x}$ and the dither $\vec{d}$.

It is assumed that $\vec{x}$ and $\vec{d}$ are independent identically distributed (IID) signals, thus their samples $x[n]$ and $d[n]$ can be modeled by the random variables $\mathbf{x}$ and $\mathbf{d}$. The characteristic function of the quantizer input $\mathbf{z}$ can be expressed by the product of the characteristic functions of $\mathbf{d}$ and $\mathbf{x}$

$$M_{\mathbf{z}}\left(ju\right) = M_{\mathbf{x}}\left(ju\right) M_{\mathbf{d}}\left(ju\right), \tag{17}$$

since the quantizer input $\mathbf{z}$ is derived by the summation of two independent random variables $\mathbf{d}$ and $\mathbf{x}$. We recall the definition of the characteristic function

$$M_{\mathbf{y}}\left(ju\right) = \mathrm{E}\left\{\mathrm{e}^{ju\mathbf{y}}\right\} = \int\limits_{-\infty}^{\infty} p_{\mathbf{y}}\left(y\right) \mathrm{e}^{juy} \;\; \mathrm{d}y = \mathcal{F}_{p_{\mathbf{y}}(y)}\left(-u\right), \tag{18}$$

where $\mathcal{F}_v\left(\omega\right)$ denotes the Fourier transform of a function v.

Schuchman [25] derived the condition

$$M_{\mathbf{d}}\left(j\frac{2\pi b}{\Delta}\right) = 0, \quad b \in \mathbb{Z},\; b \neq 0, \tag{19}$$

which is necessary and sufficient to achieve a subtractive quantization error $\mathbf{e}$ that is independent from the quantizer input $\mathbf{x}$, and an IID sequence uniformly distributed on $(-\Delta/2, \Delta/2]$. Gray and Stockham [23] showed that the condition

$$\frac{\mathrm{d}^k}{\mathrm{d}u^k}\left(M_{\mathbf{d}}\left(ju\right) M_{\mathbf{v}}\left(ju\right)\right)\Bigg|_{u=2\pi b/\Delta} = 0, \quad b \in \mathbb{Z},\ b \neq 0, \tag{20}$$

where $\mathbf{v}$ is a random variable uniformly distributed on $(-\Delta/2, \Delta/2]$ and independent of $\mathbf{d}$, is necessary and sufficient for the kth moment of the non-subtractive quantization error ϵ not to depend on $\mathbf{x}$. Note that $\mathbf{v}$ by itself has no special meaning, and is only introduced to formulate the condition (20). Both results are valid for uniform scalar quantizers with step size Δ, where overload does not occur. Schuchman's condition (19) is especially important for the analysis of quantization attacks against additive watermarks, as shown in Section 4.

3.1.2 Characteristic Function of the Quantization Error $\mathbf{e}$

The PDF of the quantization error of a uniform quantizer, having a representative value at 0, can be expressed by the sum of the PDFs of the quantization errors occurring in each bin.

$$p_{\mathbf{e}}\left(e\right) = \underbrace{\operatorname{rect}\left(\frac{e}{\Delta}\right)}_{\text{extracting the central interval}} \cdot \underbrace{\sum_{b=-\infty}^{\infty} p_{\mathbf{z}}\left(b\Delta - e\right),}_{\text{mirroring around 0 and periodic repetition}} \tag{21}$$

$$\text{where} \quad \operatorname{rect}\left(e\right) = \begin{cases} 1, & |e| < 0.5, \\ 0, & |e| \geq 0.5. \end{cases} \tag{22}$$

This description of the PDF of the quantization error can be transformed straightforwardly into an expression for the characteristic function of the error

$$M_{\mathbf{e}}\left(ju\right) = \sum_{b=-\infty}^{\infty} M_{\mathbf{z}}\left(j\frac{2\pi b}{\Delta}\right) \operatorname{si}\left(\frac{\Delta}{2}(u + 2\pi b/\Delta)\right). \tag{23}$$

The moments of a random variable $\mathbf{y}$ can be found from subsequent derivatives of its characteristic function $M_{\mathbf{y}}\left(ju\right)$,

$$\mathrm{E}\left\{\mathbf{y}^k\right\} = j^{-k} \frac{d^k}{du^k} M_{\mathbf{y}}\left(ju\right)\Bigg|_{u=0}, \tag{24}$$

leading to

$$
\begin{aligned}
\mathrm{E}\left\{\mathbf{e}\right\} &= j^{-1} \sum_{b=-\infty}^{\infty} M_{\mathbf{z}}\left(j\frac{2\pi b}{\Delta}\right) \frac{d}{du}\mathrm{si}\left(\frac{\Delta}{2}(u+2\pi b/\Delta)\right)\Bigg|_{u=0} \\
&= -j \sum_{\substack{b=-\infty \\ b\neq 0}}^{\infty} M_{\mathbf{z}}\left(j\frac{2\pi b}{\Delta}\right) \frac{(-1)^b}{2\pi b/\Delta} \qquad (25)\\
\mathrm{E}\left\{\mathbf{e}^2\right\} &= \frac{\Delta^2}{12} + \sum_{\substack{b=-\infty \\ b\neq 0}}^{\infty} M_{\mathbf{z}}\left(j\frac{2\pi b}{\Delta}\right) \frac{(-1)^b}{2(\pi b/\Delta)^2}. \qquad (26)
\end{aligned}
$$

Schuchman's condition for obtaining a quantization error $\mathbf{e}$ that is statistically independent from the input $\mathbf{x}$ can be verified by (23) and (17). If (19) is fulfilled, the values $M_{\mathbf{z}}\left(j\frac{2\pi b}{\Delta}\right)$ for $b \neq 0$ are zero, no matter what the characteristic function of the input signal $M_{\mathbf{x}}(ju)$ is.

3.2 Dependence between the Dither $\mathbf{d}$ *and the Error* $\mathbf{e}$

In Section 4 the performance of a watermark decoder after signal quantization is expressed in terms of the expectations $\mathrm{E}\left\{\mathbf{e}^2\right\}$, $\mathrm{E}\left\{\mathbf{ew}\right\}$, $\mathrm{E}\left\{\mathbf{ew}^3\right\}$ and $\mathrm{E}\left\{\mathbf{exw}^2\right\}$. In this section, expressions for these terms dependent on the characteristics of the watermark $\mathbf{w}$ and the host signal $\mathbf{x}$ are presented. The calculations are an extension of the theory of dithered scalar quantizers, thus the generality of these results will be emphasized by using the notation $\mathbf{d}$ instead of $\mathbf{w}$.

3.2.1 Notation and Normalization

First of all, some notations and normalizations are introduced. This is necessary to keep the formulas relatively small and to make results easier to compare. The integral

$$
M_{\mathbf{x}}^{(k)}(ju) = \int_{-\infty}^{\infty} x^k \, p_{\mathbf{x}}(x) \, \mathrm{e}^{jux} \, \mathrm{d}x \qquad (27)
$$

equals the k-th derivative of the characteristic function, except for a complex factor. For convenience the PDFs of the involved random variables are normalized by their standard deviation, thus we define

$$
\begin{aligned}
p_{\tilde{\mathbf{x}}}(x) &= \sigma_{\mathbf{x}}\, p_{\mathbf{x}}(\sigma_{\mathbf{x}} x) \qquad (28)\\
M_{\tilde{\mathbf{x}}}^{(k)}(ju) &= \frac{1}{\sigma_{\mathbf{x}}^k} M_{\mathbf{x}}^{(k)}(ju/\sigma_{\mathbf{x}}). \qquad (29)
\end{aligned}
$$

Normalized parameters for the standard deviation of the dither signal and the input signal are defined as

$$\zeta = \frac{\sigma_{\mathbf{d}}}{\Delta} \tag{30}$$

$$\chi = \frac{\sigma_{\mathbf{x}}}{\Delta}. \tag{31}$$

Applying this notation to (26) and normalizing by $\Delta^2/12$, the variance of the quantization noise for fine uniform scalar quantizers, yields

$$\frac{\mathrm{E}\left\{\mathbf{e}^2\right\}}{\Delta^2/12} = 1 + 12 \sum_{b=1}^{\infty} \frac{(-1)^b}{\pi^2 b^2} M_{\tilde{\mathbf{x}}}\left(j 2\pi b \chi\right) M_{\tilde{\mathbf{d}}}\left(j 2\pi b \zeta\right). \tag{32}$$

3.2.2 Signal Dependencies

Using the conditional characteristic function of the quantizer input

$$\begin{aligned} M_{\mathbf{z} \mid \mathbf{d}=d}\left(ju\right) &= \int_{-\infty}^{\infty} p_{\mathbf{z}}\left(z \mid \mathbf{z} = \mathbf{x} + \mathbf{d}; \mathbf{d} = d\right) \mathrm{e}^{juz} \;\; \mathrm{d}z \\ &= M_{\mathbf{x}}\left(ju\right) \mathrm{e}^{jud} \end{aligned} \tag{33}$$

and (25) an expression for the correlation between the quantization error $\mathbf{e}$ and the dither $\mathbf{d}$ can be derived:

$$\mathrm{E}\left\{\mathbf{e} | \mathbf{d} = d\right\} = -j \sum_{\substack{b=-\infty \\ b \neq 0}}^{\infty} \frac{(-1)^b}{2\pi b/\Delta} M_{\mathbf{x}}\left(j \frac{2\pi b}{\Delta}\right) \mathrm{e}^{j\frac{2\pi b}{\Delta} d} \tag{34}$$

$$\begin{aligned} \mathrm{E}\left\{\mathbf{ed}\right\} &= \int_{-\infty}^{\infty} d \, \mathrm{E}\left\{\mathbf{e} | \mathbf{d} = d\right\} p_{\mathbf{d}}\left(d\right) \;\; \mathrm{d}d \\ &= -j \sum_{\substack{b=-\infty \\ b \neq 0}}^{\infty} \frac{(-1)^b}{2\pi b/\Delta} M_{\mathbf{x}}\left(j \frac{2\pi b}{\Delta}\right) \int_{-\infty}^{\infty} d \, p_{\mathbf{d}}\left(d\right) \mathrm{e}^{j\frac{2\pi b}{\Delta} d} \;\; \mathrm{d}d \\ &= -j \sum_{\substack{b=-\infty \\ b \neq 0}}^{\infty} \frac{(-1)^b}{2\pi b/\Delta} M_{\mathbf{x}}\left(j \frac{2\pi b}{\Delta}\right) M_{\mathbf{d}}^{(1)}\left(j \frac{2\pi b}{\Delta}\right). \end{aligned} \tag{35}$$

Thus, the cross-correlation $\mathrm{E}\left\{\mathbf{ed}\right\}$ is dependent on the characteristic functions of the random variables $\mathbf{x}$ and $\mathbf{d}$ and the quantizer step size Δ. Assuming that the mean-free dither signal has an even symmetric PDF, the well-known symmetries of Fourier transforms of real functions can be exploited. Thus, the summation over negative b's can be eliminated. After normalizing, we obtain

$$\frac{\mathrm{E}\left\{\mathbf{ed}\right\}}{\sigma_{\mathbf{d}}^2} = \sum_{b=1}^{\infty} \frac{(-1)^b}{\pi b \zeta} M_{\tilde{\mathbf{x}}}\left(j2\pi b\chi\right) \mathcal{I}m \left\{ M_{\tilde{\mathbf{d}}}^{(1)}\left(j2\pi b\zeta\right)\right\}. \tag{36}$$

For the characterization of the quantization channel, some higher-order statistics and signal dependencies must be evaluated. Since the calculation is always similar to the presented derivation of (36), we summarize here only the resulting equations:

$$\frac{\mathrm{E}\left\{\mathbf{ex}\right\}}{\sigma_{\mathbf{x}}^2} = \sum_{b=1}^{\infty} \frac{(-1)^b}{\pi b \chi} \mathcal{I}m \left\{ M_{\tilde{\mathbf{x}}}^{(1)}\left(j2\pi b\chi\right)\right\} M_{\tilde{\mathbf{d}}}\left(j2\pi b\zeta\right) \tag{37}$$

$$\frac{\mathrm{E}\left\{\mathbf{ed}^3\right\}}{\sigma_{d}^4} = \sum_{b=1}^{\infty} \frac{(-1)^b}{\pi b \zeta} M_{\tilde{\mathbf{x}}}\left(j2\pi b\chi\right) \mathcal{I}m \left\{ M_{\tilde{\mathbf{d}}}^{(3)}\left(j2\pi b\zeta\right)\right\} \tag{38}$$

$$\frac{\mathrm{E}\left\{\mathbf{e}^2\mathbf{d}^2\right\}}{\sigma_{d}^4} = \frac{1}{12\zeta^2} + \sum_{b=1}^{\infty} \frac{(-1)^b}{\pi^2 b^2 \zeta^2} M_{\tilde{\mathbf{x}}}\left(j2\pi b\chi\right) M_{\tilde{\mathbf{d}}}^{(2)}\left(j2\pi b\zeta\right) \tag{39}$$

$$\frac{\mathrm{E}\left\{\mathbf{exd}^2\right\}}{\sigma_{d}^4} = \frac{\chi^2}{\zeta^2} \sum_{b=1}^{\infty} \frac{(-1)^b}{\pi b \chi} \mathcal{I}m \left\{ M_{\tilde{\mathbf{x}}}^{(1)}\left(j2\pi b\chi\right)\right\} M_{\tilde{\mathbf{d}}}^{(2)}\left(j2\pi b\zeta\right). \tag{40}$$

The given formulas have been verified for different signal models. Several examples will be presented in Section 4 and Section 5. For the dither signal $\mathbf{d}$ a uniform, a Gaussian and a bipolar[1] distribution are considered. Models for the input signal $\mathbf{x}$ are Gaussian, Laplacian or generalized Gaussian random variables. Except for the generalized Gaussian variable, the PDFs and the required characteristic functions are summarized in Appendix A.

4 Detection Robustness after Quantization

4.1 The Quantization Channel

In this section, the effects of subsequent uniform scalar quantization on the detection robustness of an additively embedded watermark will be analyzed. The detection performance depends on the quantizer step size Δ and on the statistical properties of the watermark $\vec{w}$ and the host signal $\vec{x}$. The considered scheme is depicted in Fig. 3. Considering only the quantization attack, the channel distortion $\vec{e}$ is equal to the quantization error $e[n] = r[n] - s[n]$.

[1] $d \in \{-\sigma_{\mathbf{d}}, \sigma_{\mathbf{d}}\}$, where both signs are equi-probable

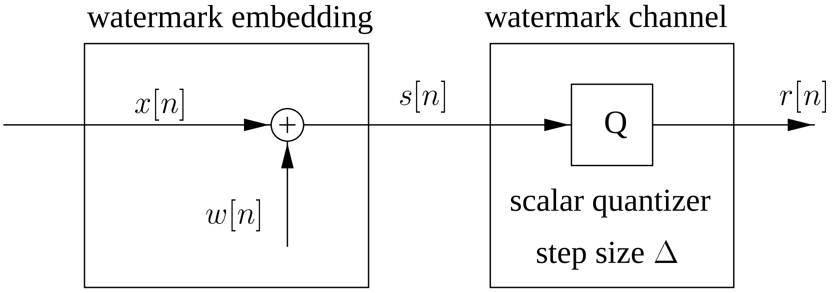

Fig. 3. Additive watermark embedding and quantization attack

Characterizing the quantization channel means describing the dependency between the received quantized signal $\vec{r}$, the original signal $\vec{x}$, and the watermark $\vec{w}$. We assume that the watermark $\vec{w}$ and the host signal $\vec{x}$ are independent. The samples of both signals are independent identically distributed (IID), and drawn from wide-sense stationary random processes denoted by the random variables $\mathbf{w}$ and $\mathbf{x}$. For convenience, we again use the normalized standard deviations $\chi = \sigma_{\mathbf{x}}/\Delta$ and $\zeta = \sigma_{\mathbf{w}}/\Delta$. The watermark-to-host-document ratio (WDR), defined by $\mathrm{WDR} = 20\log_{10}(\zeta/\chi) = 20\log_{10}(\sigma_w/\sigma_x)$ determines the amplitude of the watermarks. Small WDRs correspond to a high perceptual quality since in this case the host document is not distorted very much by the watermarking process.

4.2 Infinite Length Watermarks

Using the correlation detection principle presented in Section 2, the detection robustness can be determined when $\mathrm{E}\left\{\mathbf{c}|\mathrm{H}_1\right\}$, Var $\left\{\mathbf{c}|\mathrm{H}_1\right\}$ and Var $\left\{\mathbf{c}|\mathrm{H}_0\right\}$ are known. For the scheme depicted in Fig. 3, the normalized conditional expectation $\mathrm{E}\left\{\mathbf{c}|\mathrm{H}_1\right\}$ is given by

$$\begin{aligned}\frac{\mathrm{E}\left\{\mathbf{c}|\mathrm{H}_1\right\}}{\sigma_{\mathbf{w}}^2} = \frac{\mathrm{E}\left\{(\mathbf{r}-\gamma_x\mathbf{x})\mathbf{w}\right\}}{\sigma_{\mathbf{w}}^2} &= \frac{\mathrm{E}\left\{(\mathbf{e}+\mathbf{w}+(1-\gamma_x)\mathbf{x})\mathbf{w}\right\}}{\sigma_{\mathbf{w}}^2}\\ &= \frac{\mathrm{E}\left\{\mathbf{ew}\right\}}{\sigma_{\mathbf{w}}^2} + 1, \end{aligned} \tag{41}$$

where the received signal $\mathbf{r}$ was replaced using (2). Further, the independence of $\mathbf{w}$ and $\mathbf{x}$ was exploited, and we assumed $\mathrm{E}\left\{\mathbf{w}\right\} = 0$. In the ideal case we would like to obtain $\mathrm{E}\left\{\mathbf{c}|\mathrm{H}_1\right\}/\sigma_{\mathbf{w}}^2 = 1$, i.e., the correlation $\mathrm{E}\left\{\mathbf{ew}\right\}$ between the quantization error $\mathbf{e}$ and the watermark $\mathbf{w}$ should be zero.

According to (10), the influence of the variances Var $\left\{\mathbf{c}|\mathrm{H}_1\right\}$ and Var $\left\{\mathbf{c}|\mathrm{H}_0\right\}$ on the detection robustness decreases for an increasing correlation length M. Therefore, it is possible to describe the detection of infinite length watermarks completely by the expectation $\mathrm{E}\left\{\mathbf{c}|\mathrm{H}_1\right\}$ given in (41). Whether or not $\mathrm{E}\left\{\mathbf{ew}\right\}$ is zero, depends on the PDFs of the original signal $\mathbf{x}$ and the watermark $\mathbf{w}$ and their strength relative to the quantizer step size Δ. This can be deduced from (36), when substituting the dither $\mathbf{d}$ by the watermark $\mathbf{w}$.

Schuchman's condition (19) is necessary and sufficient to make the input signal $\mathbf{x}$ and the quantization error $\mathbf{e}$ independent. Since this solution is symmetric with respect to the original and the dither signal we obtain the condition

$$M_{\mathbf{x}}\left(j\frac{2\pi b}{\Delta}\right) = 0, \quad b \in \mathbb{Z},\ b \neq 0 \tag{42}$$

on the characteristic function of the input signal to achieve $\mathrm{E}\left\{\mathbf{ew}\right\} = 0$ for any watermark. (42) is fulfilled when the PDF of $\mathbf{x}$ can be expressed by the convolution of the PDF of any random variable with a uniform distribution on $(-\Delta/2; \Delta/2]$. This condition is approximately fulfilled for PDFs which are almost constant over the range of one quantization bin, which is typically achieved by fine quantization.

In conjunction with signal transforms and coarse quantization, as typical for lossy compression systems, the distribution of the input signal often does not fulfill (42). Therefore, we evaluated (41) using (36) for different assumptions about the input signal, the watermark (dither), and the quantizer step size. We are mainly interested in the robustness of a specific additive watermark against differently coarse quantization. Thus (41) must be evaluated for a constant WDR, describing the embedding strength, and increasing quantizer step size Δ, which equals in this context a decreasing value of χ, the normalized standard deviation of the host document.

Fig. 4 and Fig. 5 show the cross-correlation $\mathrm{E}\left\{\mathbf{yw}\right\}$ of the watermark $\mathbf{w}$ and the pre-processed received signal $\mathbf{y}$, where the curves are computed using (36). We consider Gaussian and Laplacian host signals $\mathbf{x}$ with zero mean and unit variance. The watermark distribution is either Gaussian, uniform or bipolar ($w[n] = \pm\sigma_{\mathbf{w}}$). All three types of signals are frequently used in watermarking schemes. Results for $\mathrm{WDR} = -16.59\mathrm{dB}$ and $\mathrm{WDR} = -6.02\mathrm{dB}$ are presented. The cross-correlation $\mathrm{E}\left\{\mathbf{yw}\right\}$ is plotted over varying values of χ in the upper plots of Fig. 4 and Fig. 5. These plots show the quantization effects dependent on the quantizer step size. In the lower plots, the cross-correlation $\mathrm{E}\left\{\mathbf{yw}\right\}$ is plotted over the host-document-to-noise ratio (DNR), defined by $\mathrm{DNR} = 10\log_{10}(\sigma_{\mathbf{x}}^2/\mathrm{Var}\left\{\mathbf{x}-\mathbf{r}\right\})$ and denoting the distortion between the attacked document $\mathbf{r}$ and the unwatermarked original document $\mathbf{x}$. In contrast to the upper plots, the curves in the lower plots are additionally influenced by the dependence of the quantization noise power on the PDF of the signal. Fig. 5 shows that it is possible to obtain even negative DNR-values. This effect occurs since the DNR includes the distortion introduced by the watermark.

For fixed WDR and varying χ, the cross-correlation $\mathrm{E}\left\{\mathbf{yw}\right\}$ becomes one for sufficiently large χ (fine quantization) and converges towards zero in the limit as $\chi \to 0$. The behavior of $\mathrm{E}\left\{\mathbf{yw}\right\}$ for large χ is intuitively clear, since in this case the host signal has an approximately constant PDF over the range of a step size Δ, and thus

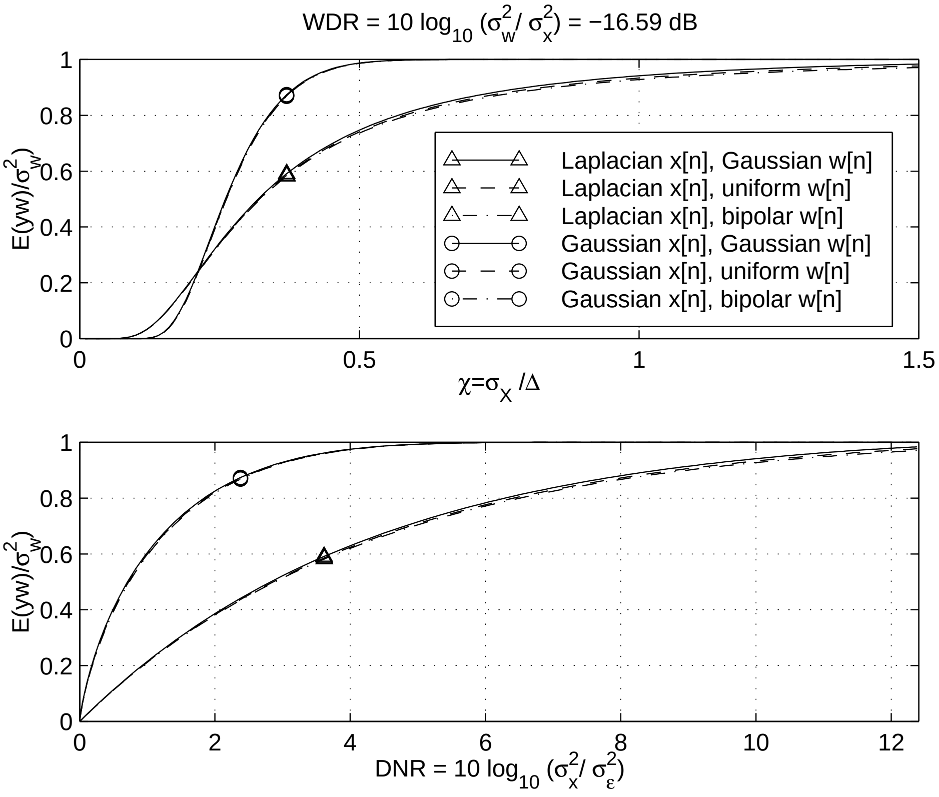

Fig. 4. Cross-correlation $\mathrm{E}\left\{\mathbf{yw}\right\}/\sigma_{\mathbf{w}}^2$ for $\mathrm{WDR} = -16.59\mathrm{dB}$.

(42) is approximately fulfilled. At the limit $\chi \to 0$ the quantizer step size Δ becomes arbitrarily large, which leads to a zero quantizer output, assuming zero is a representation value, and thus to the quantization error $e[n] = -x[n] - w[n]$. As a result, the normalized expectation $\mathrm{E}\left\{\mathbf{ew}\right\}/\sigma_{\mathbf{w}}^2$ converges to -1, and the conditional expectation $\mathrm{E}\left\{\mathbf{c}|\mathrm{H}_1\right\}$ becomes zero.

Let us now compare the results for different watermark distributions. We observe that the characteristic of the watermark PDF does not have a significant influence for $\mathrm{WDR} = -16.59\mathrm{dB}$ (low power embedding). However, for larger WDRs, as shown in Fig. 5, a Gaussian watermark provides a somewhat better robustness against quantization. For instance, for a Laplacian distributed host signal and $\mathrm{WDR} = -6.02\mathrm{dB}$, the quality after a quantization attack that reduces the normalized cross-correlation $\mathrm{E}\left\{\mathbf{yw}\right\}/\sigma_{\mathbf{w}}^2$ to 0.8 is about 1 dB higher for a bipolar distributed watermark compared to a Gaussian watermark. Thus, the bipolar distributed watermark can be erased more easily by quantization.

The shape of the host signal distribution is much more important than the watermark distribution. The upper plots in Fig. 4 and Fig. 5 show that in most cases a greater cross-correlation $\mathrm{E}\left\{\mathbf{yw}\right\}$ is preserved for lower values of χ in the case of a Gaussian distributed host signal. Only for very coarse quantization (very small χ), the Laplacian distributed host offers better robustness against quantization. However, the quantizer step size Δ does not directly indicate the quality of the attacked

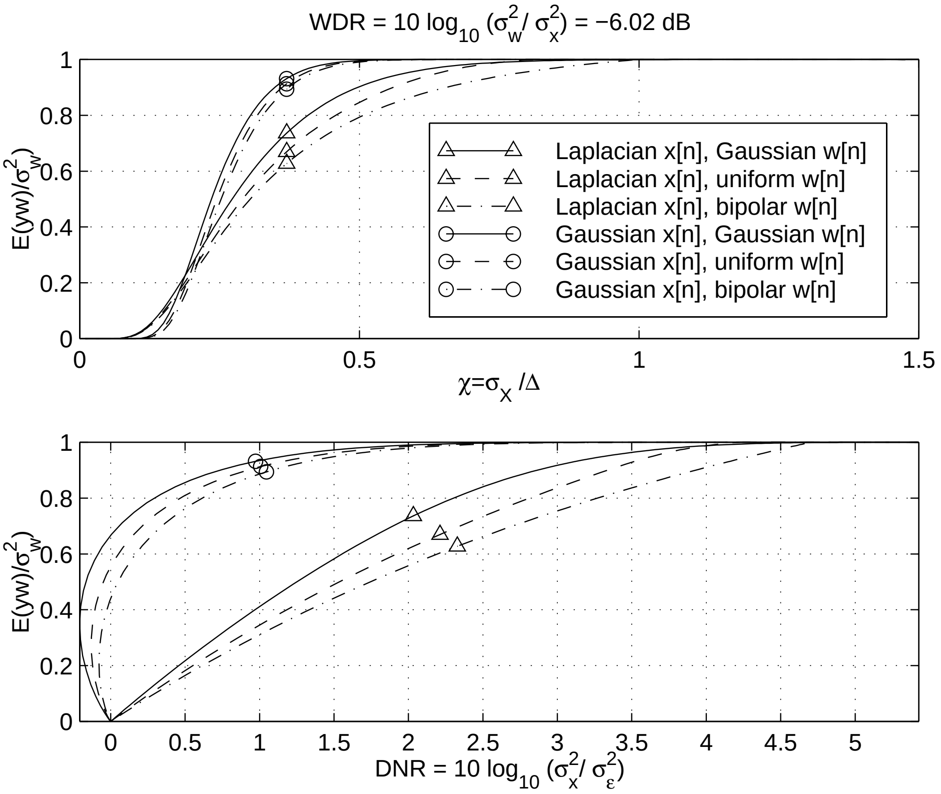

Fig. 5. Cross-correlation $\mathrm{E}\left\{\mathbf{yw}\right\}/\sigma_{\mathbf{w}}^2$ for $\mathrm{WDR} = -6.02\mathrm{dB}$.

document. From the lower plots in Fig. 4 and Fig. 5, where the cross-correlation is plotted over the DNR, we conclude that the watermark with a Gaussian distributed host resists quantization more strongly than an equivalent watermark in a Laplacian distributed host for all qualities of the attacked document. In the case of a Gaussian distributed host, $\mathrm{E}\left\{\mathbf{yw}\right\}$ drops to zero at much lower DNRs. In general, we observe that with more peaky PDFs, everything else being equal, the watermark is somewhat less robust against quantization attacks. This becomes important in Section 5 when the watermarking of natural image data is investigated. Further, due to the strong dependency of the cross-correlation $\mathrm{E}\left\{\mathbf{yw}\right\}$ on the quantization strength for peaky host document PDFs, it is not possible to model the quantization channel by an AWGN channel.

4.3 Robustness of Finite Length Watermarks

In the previous sub-section, the expectation $\mathrm{E}\left\{\mathbf{c}|\mathrm{H}_1\right\}$ was investigated, since this term describes completely the robustness of infinite length watermarks. However, in practice we have finite-length signals, where the detection performance depends strongly on the variances $\mathrm{Var}\left\{\mathbf{c}|\mathrm{H}_1\right\}$ and $\mathrm{Var}\left\{\mathbf{c}|\mathrm{H}_0\right\}$. In addition, the availability of the original document to the watermark detector becomes an important factor for the watermark robustness. The detection robustness can always be improved by an increased correlation length M. In the following comparisons the correlation length is set to $M = 10000$, which gives realistic error probabilities, particularly for the worst case (blind detection).

4.3.1 Analytic Expressions for Var $\{\mathbf{c}|\mathrm{H}_1\}$ *and Var* $\{\mathbf{c}|\mathrm{H}_0\}$

The formulas for the variances Var $\{\mathbf{c}|\mathrm{H}_1\}$ and Var $\{\mathbf{c}|\mathrm{H}_0\}$ are slightly more complicated than those for $\mathrm{E}\{\mathbf{c}|\mathrm{H}_1\}$. Nevertheless, they can be derived similarly as (41). In the case of hypothesis H_1 the received signal $\mathbf{r}$ must be replaced by (2). For hypothesis H_0, two formulas are derived. (44) is valid when considering the received signal $\mathbf{r} = \mathbf{x} + \mathbf{e}$. This means the watermark test is conducted for a document that is not watermarked at all. However, it could happen that an embedded watermark $\mathbf{w_1}$ with $\mathrm{E}\{\mathbf{w_1 w}\} = 0$ exists. Under the assumption that the watermark $\mathbf{w_1}$ has the same distribution – including the variance – as the tested watermark $\mathbf{w}$, (45) is valid. (45) can be obtained by setting $\mathbf{r} = \mathbf{x}+\mathbf{e}+\mathbf{w_1}$ for the received signal. The attack in the cases H_0 and H_0^* is equal to the quantization attack in case of H_1. Usually, we cannot expect that the input signal has zero mean. This is considered in (43), (44) and (45) by the term $\mathrm{E}\{\mathbf{x}\}$.

$$\begin{aligned}
\mathrm{H}_1:\ \frac{\text{Var}\{(\mathbf{r}-\gamma_x\mathbf{x})\mathbf{w}\}}{\sigma_{\mathbf{w}}^4} &= \frac{\mathrm{E}\{\mathbf{e}^2\mathbf{w}^2\}}{\sigma_{\mathbf{w}}^4} + 2\frac{\mathrm{E}\{\mathbf{e}\mathbf{w}^3\}}{\sigma_{\mathbf{w}}^4} + \frac{\mathrm{E}\{\mathbf{w}^4\}}{\sigma_{\mathbf{w}}^4} \\
&\quad + 2(1-\gamma_x)\frac{\mathrm{E}\{\mathbf{e}\mathbf{x}\mathbf{w}^2\}}{\sigma_{\mathbf{w}}^4} - \left(\frac{\mathrm{E}\{\mathbf{e}\mathbf{w}\}}{\sigma_{\mathbf{w}}^2} + 1\right)^2 \\
&\quad + (1-\gamma_x)^2\left(\frac{\chi^2}{\zeta^2} + \frac{\mathrm{E}\{\mathbf{x}\}^2}{\sigma_{\mathbf{w}}^2}\right) \qquad (43) \\
\mathrm{H}_0:\ \frac{\text{Var}\{(\mathbf{r}-\gamma_x\mathbf{x})\mathbf{w}\}}{\sigma_{\mathbf{w}}^4} &= \frac{1}{12\zeta^2}\frac{\mathrm{E}\{\mathbf{e}^2\}}{\Delta^2/12} + 2(1-\gamma_x)\frac{\chi^2}{\zeta^2}\frac{\mathrm{E}\{\mathbf{e}\mathbf{x}\}}{\sigma_{\mathbf{x}}^2} \\
&\quad + (1-\gamma_x)^2\left(\frac{\chi^2}{\zeta^2} + \frac{\mathrm{E}\{\mathbf{x}\}^2}{\sigma_{\mathbf{w}}^2}\right) \qquad (44) \\
\mathrm{H}_0^*:\ \frac{\text{Var}\{(\mathbf{r}-\gamma_x\mathbf{x})\mathbf{w}\}}{\sigma_{\mathbf{w}}^4} &= 1 + \frac{1}{12\zeta^2}\frac{\mathrm{E}\{\mathbf{e}^2\}}{\Delta^2/12} + 2(1-\gamma_x)\frac{\chi^2}{\zeta^2}\frac{\mathrm{E}\{\mathbf{e}\mathbf{x}\}}{\sigma_{\mathbf{x}}^2} \\
&\quad + 2\frac{\mathrm{E}\{\mathbf{e}\mathbf{w}\}}{\sigma_{\mathbf{w}}^2} + (1-\gamma_x)^2\left(\frac{\chi^2}{\zeta^2} + \frac{\mathrm{E}\{\mathbf{x}\}^2}{\sigma_{\mathbf{w}}^2}\right). \qquad (45)
\end{aligned}$$

These equations show that the variances Var $\{\mathbf{c}|\mathrm{H}_1\}$ and Var $\{\mathbf{c}|\mathrm{H}_0\}$ are not exactly equal. Later we will see whether this difference is important in practical applications.

As mentioned in Section 1, blind detection denotes the case of $\gamma_x = 0$. When the original $\mathbf{x}$ is known, the optimal weight, removing all the interference from $\mathbf{x}$, can be determined by the correlation of the received signal $\mathbf{r}$ with the original signal $\mathbf{x}$:

$$\gamma_x = \frac{\mathrm{E}\left\{\mathbf{rx}\right\}}{\sigma_{\mathbf{x}}^2} = \frac{\mathrm{E}\left\{\mathbf{ex}\right\}}{\sigma_{\mathbf{x}}^2} + \frac{\mathrm{E}\left\{\mathbf{x}\right\}^2}{\sigma_{\mathbf{x}}^2} + 1. \tag{46}$$

4.3.2 Discussion of the Results

In the following discussion, the WDR is fixed to $-16.59\mathrm{dB}$. In Section 4.2 it is shown that for small WDR the watermark distribution has no significant influence on $\mathrm{E}\left\{\mathbf{c}|\mathrm{H}_1\right\}$. Similar results are obtained for Var $\left\{\mathbf{c}|\mathrm{H}_1\right\}$ and Var $\left\{\mathbf{c}|\mathrm{H}_0\right\}$. Therefore, we will present only results for a Gaussian watermark. For the hypothesis H_0, a completely unmarked document is assumed, thus (44) has to be used to compute Var $\left\{\mathbf{c}|\mathrm{H}_0\right\}$.

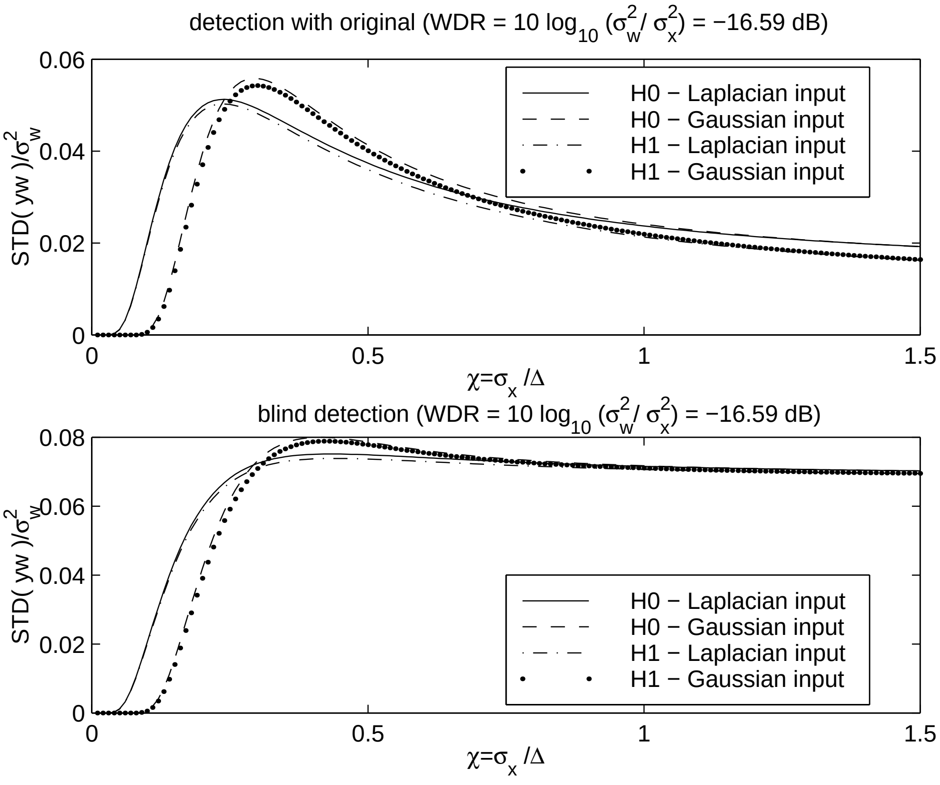

Fig. 6. Standard deviations $\sqrt{\text{Var}\left\{\mathbf{yw}\right\}}/\sigma_{\mathbf{w}}^2$ for $\mathrm{WDR} = -16.59\mathrm{dB}$.

Fig. 6 shows the normalized standard deviation $\sqrt{\text{Var}\left\{\mathbf{yw}\right\}}/\sigma_{\mathbf{w}}^2$ for Laplacian and Gaussian input sequences over the quantization strength denoted by χ. Detection with original and blind detection is considered. The values for both hypotheses H_0 and H_1 are displayed in one plot. We observe that the standard deviations for small χ and the same input characteristic are almost identical for H_0 and H_1. This is no longer true for large χ (fine quantization), as can easily be seen in the case of detection with original. The same effect exists also for blind detection, but it is less prominent due to the large interference from the original signal. The variances for large χ can be estimated by assuming that the quantization error e becomes statistically independent from the quantizer input signal (Schuchman's condition is approximately fulfilled). Especially when detecting with the original, the equations for the variances simplify considerably, to

$$\mathrm{H}_1: \ \frac{\text{Var}\left\{(\mathbf{r}-\gamma_x\mathbf{x})\mathbf{w}\right\}}{\sigma_{\mathbf{w}}^4} = \frac{1}{12\zeta^2} + \frac{\mathrm{E}\left\{\mathbf{w}^4\right\}}{\sigma_{\mathbf{w}}^4} - 1 \tag{47}$$

$$\mathrm{H}_0: \ \frac{\text{Var}\left\{(\mathbf{r}-\gamma_x\mathbf{x})\mathbf{w}\right\}}{\sigma_{\mathbf{w}}^4} = \frac{1}{12\zeta^2} \tag{48}$$

$$\mathrm{H}_0^*: \ \frac{\text{Var}\left\{(\mathbf{r}-\gamma_x\mathbf{x})\mathbf{w}\right\}}{\sigma_{\mathbf{w}}^4} = \frac{1}{12\zeta^2} + 1 \tag{49}$$

The first equation reveals that the choice of the watermark distribution becomes important for large values of ζ. For instance, the kurtosis $\mathrm{E}\left\{\mathbf{w}^4\right\}/\sigma_{\mathbf{w}}^4$ is 1 for a bipolar watermark and 3 for a Gaussian watermark. Since for fine quantization the correlation mean $\mathrm{E}\left\{\mathbf{c}|\mathrm{H}_1\right\}$ is not affected, here bipolar watermarks are superior to Gaussian watermarks. Note that for coarse quantization the opposite is true, due to the decreasing cross-correlation $\mathrm{E}\left\{\mathbf{yw}\right\}$.

Another effect visible in Fig. 6 is the larger standard deviation STD$\left\{\mathbf{yw}\right\}$ for Laplacian input when small χ are considered. On the other hand, for intermediate quantizer step sizes the Gaussian input leads to a larger standard deviation and for fine quantization (large χ) the distribution of the input signal does not play a significant role. Further, it can be seen that for coarse quantization the standard deviations do not depend much on the availability of the original. In this case the quantization noise dominates the distortion.

4.3.3 Error Probabilities after Quantization Attacks

Finally, the detection error probabilities can be computed from $\mathrm{E}\left\{\mathbf{c}|\mathrm{H}_1\right\}$, Var$\left\{\mathbf{c}|\mathrm{H}_1\right\}$ and Var$\left\{\mathbf{c}|\mathrm{H}_0\right\}$, as described in Section 2. Fig. 7 shows the error probabilities that occur after quantization attacks for the same parameter settings as in Fig. 6. Significant differences are observed for Laplacian and Gaussian input signals. Mainly due to the faster decay of $\mathrm{E}\left\{\mathbf{c}|\mathrm{H}_1\right\}$ for Laplacian signals, these kind of host signals provide less robustness for watermarking than Gaussian signals. For instance, error probabilities about 10^{-5} for blind detection in the case of a Laplacian host signal can be achieved via quantization introducing 3 dB lower distortion than for Gaussian host signals. This effect is even stronger for PDFs that are more peaky than the Laplacian PDF (see also Section 5). Further, we observe that detection with the original is hardly affected by fine quantization. The given plot is truncated at $p_e = 10^{-30}$. In the case of blind detection the achievable error rate is limited by the interference from the host signal.

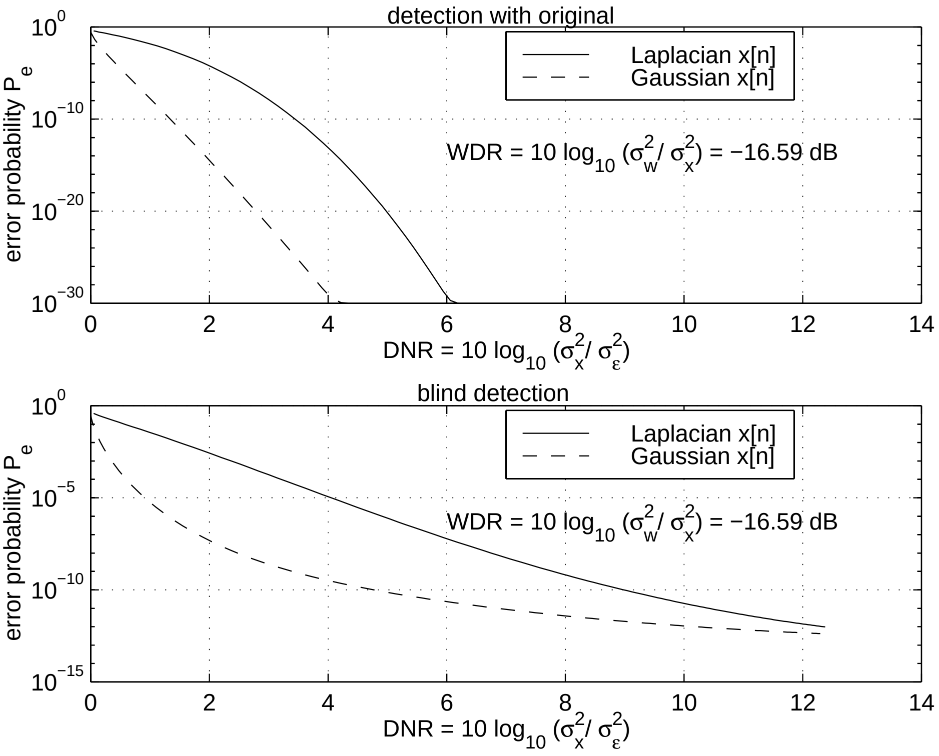

Fig. 7. Error probabilities for $\mathrm{WDR} = -16.59\mathrm{dB}$.

5 Robustness of Additive Image Watermarks to JPEG Compression

The theoretical analysis presented in the previous section shows that quantization can remove at least parts of an embedded additive watermark. This effect is dependent on the distribution of the host signal $\vec{x}$, the watermark signal $\vec{w}$ and the quantization step size Δ. The results presented so far are simply parameterized by ζ and χ, describing the watermark's and input signal's standard deviation relative to the quantizer step size Δ. However, we do not know yet which settings for ζ and χ are likely to appear in a practical environment. More insight can be gained by an example watermarking scheme. Further, the experiments can show how accurate the statistical model for the host data must be to achieve a sufficiently good prediction of the expected detection error probability after quantization attacks.

5.1 Host Data

The theoretical analysis has been carried out without specifying the data to be watermarked, so the results can be applied to many different signals. The following experimental investigations are for natural images. The watermark is embedded into the coefficients of an 8×8 block-DCT of the luminance component. Many different domains for the watermark embedding process have been proposed in recent publications, where besides the DCT domain, wavelet domains are very popular [9,11,26,27,16]. We choose the blockwise DCT transform since it is also used within the popular JPEG compression scheme.

For JPEG compression, the coefficients of the 8×8 block-DCT are quantized with a uniform scalar quantizer, as treated in the presented theory. The coefficients of the 8×8 block-DCT are almost uncorrelated. For common images, the dependencies between coefficients of the same frequency from different blocks are low. These dependencies are small enough that the coefficients can be approximately described as IID host signals, which was assumed in the presented theory.

5.2 Embedding Scheme

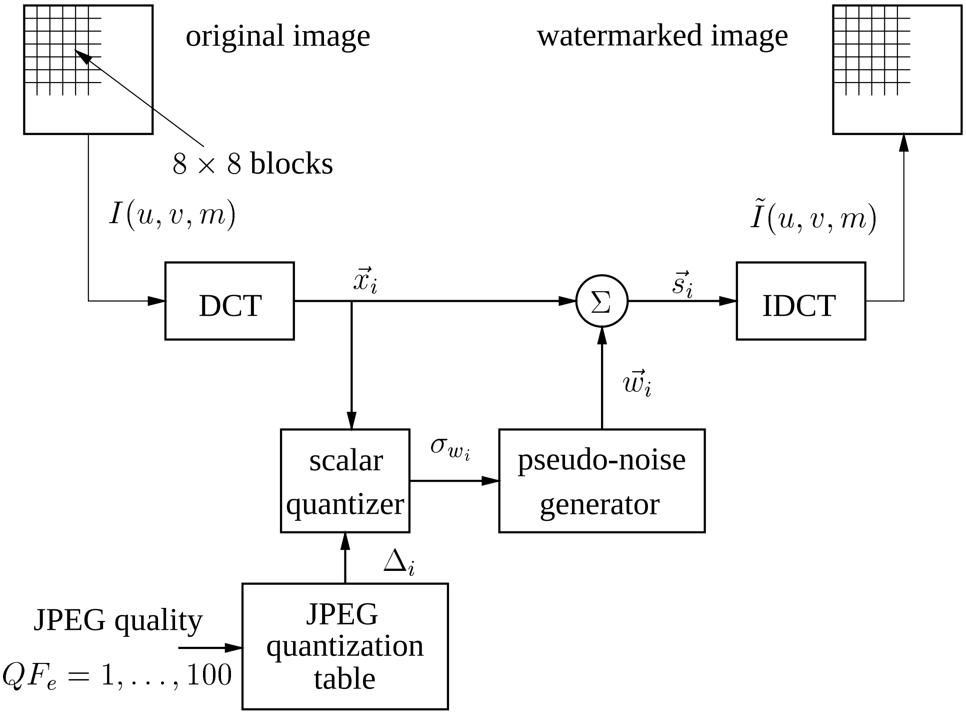

Fig. 8. Additive watermarking of DCT-coefficients

Fig. 8 depicts the scheme for the signal dependent additive watermark embedding. The signal decomposition is closely related to JPEG compression [28]. Image samples are denoted by $I(u, v, m)$, where (u, v) are the row and column indices of the mth block (where the blocks are numbered in row scan). All blocks are DCT transformed, and the coefficients for the same frequency from all blocks are grouped into a sample sequence - a *sub-signal*. Due to the 8×8 blocks, this scheme gives 64 vectors $\vec{x}_i$, where the index i denotes the sub-signal number. A different watermark sub-signal $\vec{w}_i$ is embedded in each sub-signal. Since the JPEG quantizer step sizes Δ_i are different for all frequencies, 64 different *sub-channels* have to be considered. The sub-channels are numbered according to the common zigzag scan of the DCT coefficients. Throughout the paper, no channel coding is considered explicitly. However, we investigate the watermark correlation measured per sub-channel. This approach can be considered as channel coding using a repetition code of length M_i. The resulting error rates can be translated into equivalent results for more sophisticated channel codes. Here, the length M_i of the vectors $\vec{x}_i$ equals the number of 8×8 blocks in the given image.

The quantizer step sizes for JPEG-baseline compression are optimized for subjective quality and can be parameterized via a scalar quality factor[2] QF. The main idea for the adaptation of the watermark strength is that the embedding should introduce roughly the same distortion as JPEG compression with a certain quality factor QF_e (index e for 'embedding'). This way, we achieve a quantifiable subjective quality without conducting extensive subjective tests. To realize this approach, uniform scalar quantization with step size Δ_i (as it is used in JPEG compression with a certain quality QF_e) is applied to the elements of the corresponding vector $\vec{x}_i$. The variance of the introduced distortion is measured. Estimating the quantization noise via the high-resolution result $\Delta_i^2/12$ would be not appropriate, since the actual distortion is especially for the strongly quantized high frequency components much lower than $\Delta_i^2/12$.

The watermark variance $\sigma_{w_i}^2$ in every sub-channel is chosen equal to the measured power of the quantization noise. This embedding scheme does not fulfill the power spectrum condition derived in [29,30]. However, we found that the presented scheme gives better perceptual quality than methods strictly following the power spectrum condition. After all $\sigma_{w_i}^2$ are determined, a Gaussian pseudo-noise vector $\vec{w}_i$ with the corresponding standard deviation is generated for each sub-channel and added to $\vec{x}_i$. The seed for the pseudo-noise generator can be considered the secure key for the watermarking scheme. Finally, the elements of the resulting 64 watermarked vectors $\vec{s}_i$ are transformed by an inverse DCT.

5.3 Statistical Models for DCT Coefficients

In Section 4 only Gaussian and Laplacian random variables $\mathbf{x}$ were considered as models for the host data. However, it is well known [31,32] that DCT coefficients of images can be modeled more accurately via a generalized Gaussian random variable. The PDF of a generalized Gaussian random variable with mean μ, standard deviation σ, and shape factor ν is given by

$$p_{\mathbf{x}}(x) = \frac{\nu}{2\Gamma(1/\nu)} b(\sigma, \nu)\, \mathrm{e}^{-b(\sigma,\nu)^{\nu}\,|x-\mu|^{\nu}}, \tag{50}$$

[2] $QF = 100$, highest quality with step size $\Delta = \Delta_{\min}$ for all coefficients; $QF = 1$: lowest quality with step size $\Delta = 256\Delta_{\min}$

where

$$b(\sigma,\nu) = \frac{1}{\sigma}\sqrt{\frac{\Gamma(3/\nu)}{\Gamma(1/\nu)}} \text{ and } \Gamma(a) = \int_{0}^{\infty} u^{a-1}\,\mathrm{e}^{-u}\;\mathrm{d}u. \tag{51}$$

The PDF given in (50) reduces to that of a Gaussian or Laplacian random variable for $\nu = 2$ or $\nu = 1$, respectively. Methods for estimating the parameter of the generalized Gaussian model from sample data are described in [31,33]. Since we are not aware of a closed-form expression for the characteristic function of the PDF given in (50), the samples of the characteristic function are computed numerically where necessary.

Besides the described statistical models, the specific PDF of the host sub-signal $\vec{x}_i$ is estimated. For this estimated PDF, the characteristic function is derived numerically, too. In the case of short signals, the estimated PDF cannot be considered as a statistical model, rather the results are specific to the realization of the given host signals $\vec{x}_i$.

5.4 Simulation Settings

In order to reduce the number of free parameters, we will discuss only the results for an embedding quality of $QF_e = 70$, which gives a watermarked image with sufficiently high quality. As a test image, we use the 256×256 gray-scale “Lenna” picture. The given image size leads to 1024 8×8 blocks and, thus to $M_i = 1024$ samples for each sub-channel $\vec{x}_i$.

200 differently watermarked images $\tilde{I}(u,v,m)$ were produced, using the scheme depicted in Fig. 8, where the different watermarks were obtained by different seeds for the pseudo-random number generator. The watermarked images were JPEG compressed and decompressed, each with 20 different quality factors QF_a (index a for ’attack’) between $QF_a = 5$ and $QF_a = 100$.

For watermark detection, the attacked public document is transformed again by the 8×8 block-DCT. Then the signals $\vec{y}_i$ for the different sub-channels i are correlated with the corresponding watermarks $\vec{w}_i$. Here, we will only discuss the detection performance for single sub-channels. In practical schemes, the detection results from all sub-channels can be combined to achieve a maximum robust watermark detection, as mentioned in Section 1.

For a fair test, the detection process is carried out for both hypotheses H_1 and H_0, i.e., for documents that are or are not watermarked by $\vec{w}$. For simplicity we choose as reference a completely un-watermarked image. Therefore, the resulting detection variance can be computed theoretically with (44). The un-watermarked image is always compressed in the same way that the watermarked image was compressed.

5.5 Discussion of Experimental Results

In general, we found that the experimentally derived results match the results predicted by the theory presented in Section 4. Naturally, the best match can be achieved when using the estimated PDF to describe the host sub-signals $\vec{x}_i$. The generalized Gaussian model usually delivers similar results, but the Laplacian or Gaussian models fit only in some few cases. We will briefly discuss the measured parameters like χ, DNR per sub-channel, embedding quality WDR and the shape parameter of the generalized Gaussian model. Then the expected watermark correlation after quantization will be discussed for different sub-channels. The length of the sub-channels becomes important when regarding the standard deviation of the measured correlation values. In this case, we also have to distinguish between detection with original and blind detection. Finally, measured detection error probabilities are compared with those derived theoretically.

5.5.1 Characteristics of the Test Data

Due to space constraints, it is not possible to describe and discuss the results for all different simulation parameters. Therefore, some representative results are selected. Here, we discuss only results for the sub-channels 1, 10 and 22. The detection results are presented parameterized by the quality factor QF_a of the JPEG compression attack. Table 1 shows the ratios $\chi = \sigma_{\mathbf{x}}/\Delta$ and the corresponding host-document-to-noise ratio (DNR) measured after JPEG compression for different quality factors QF_a. Thus, it is possible to relate the presented results to those from Section 4. The last line in Table 1 shows the watermark-to-host-document ratio (WDR), which characterizes the embedding strength. The values reveal that the chosen embedding scheme leads to much stronger embedding distortion (higher WDR) in the high frequency sub-channels. The WDR of the sub-channel 10 is approximately equal to the WDR considered in Section 4. Note that the DNR for sub-channel 22 does not increase monotonically with increasing JPEG quality factor QF_a. This effect has been observed for all sub-channels with large WDR. Since the distortion is measured relative to the original coefficients, this distortion measurement includes watermark noise and quantization noise. However, quantization does also reduce the watermark noise to some extent, which leads to the observed effect.

The parameter of the generalized Gaussian models for the considered sub-channels are given in Table 2. In JPEG compression the image samples are shifted from unsigned integers to signed integers [28], thus negative samples of the sub-channel

Table 1
Strength of quantization attack and distortion dependent on the quality factor QF_a. The first three columns show the ratio $\chi = \sigma_{\mathbf{x}}/\Delta$. The last three columns show the sub-channel DNR measured after JPEG attacks.

	χ			DNR per sub-channel in dB		
QF_a	$i = 1$	$i = 10$	$i = 22$	$i = 1$	$i = 10$	$i = 22$
10	4.63	0.22	0.02	23.93	3.02	-0.05
20	9.26	0.44	0.04	29.86	6.45	-0.01
30	13.72	0.66	0.06	32.87	8.81	-0.01
40	18.52	0.85	0.08	35.25	10.11	0.09
50	23.15	1.09	0.10	36.54	11.37	0.34
60	28.49	1.39	0.13	37.85	12.26	0.48
70	37.04	1.91	0.17	39.00	13.48	0.48
80	61.74	2.54	0.25	40.69	14.50	-0.07
90	123.48	5.08	0.49	41.62	15.84	-0.20
100	370.43	15.25	4.90	41.97	16.44	1.40
Embedding Quality (WDR/dB)				-42.075	-16.622	-1.527

1 (DC DCT-coefficients) can occur. Further, a factor of 8 is introduced in each DCT-coefficient due to the fast algorithm for the DCT. The samples of the sub-channel 1 are almost uniformly distributed. Therefore, the shape factor $\nu_{\mathbf{x}_1}$ is relatively large. For all other sub-channels, shape factors $\nu_{\mathbf{x}_i} < 1$ have been measured. Thus, the AC DCT-coefficients have a distribution that is more peaky than that of a Laplacian random variable.

Table 2
Parameter of generalized Gaussian model for the considered sub-channels.

sub-channel	$\mu_{\mathbf{x}}$	$\sigma_{\mathbf{x}}$	$\nu_{\mathbf{x}}$
$i = 1$	-1876.532	2963.445	3.136
$i = 10$	0.409	122.007	0.530
$i = 22$	-1.851	39.232	0.684

5.5.2 Expected Watermark Correlation per Sub-Channel

Table 1 reveals that for sub-channel 1, corresponding to the DC values of all blocks, JPEG compression applies relatively fine quantization. Even for $QF_a = 10$, the normalized parameter χ is still much larger than 1. In addition, the flat distribution of the samples of sub-channel 1 allows for the usage of fine quantization theory. Following the arguments given in Section 4.2, fine quantization does not decrease the watermark correlation. Our experiments confirm this result. We do not present a plot of the measured cross-correlation since it is almost constant for all values of QF_a.

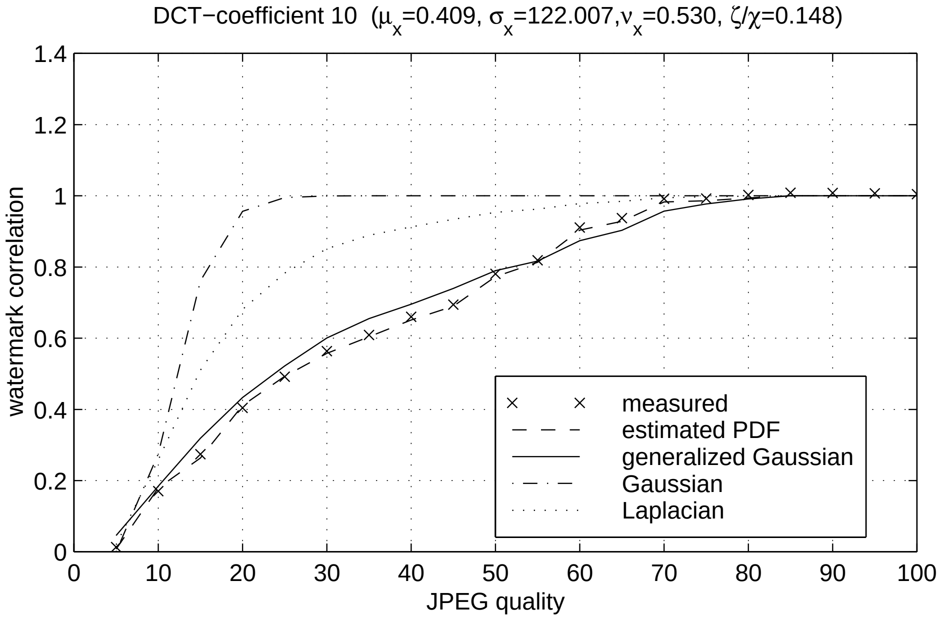

Fig. 9. Watermark correlation for sub-channel 10

Fig. 9 depicts the cross-correlation $\mathrm{E}\left\{\mathbf{y}\mathbf{w}|\mathrm{H}_1\right\}$ for sub-channel 10. We observe that the results found by the estimated PDF and the generalized Gaussian model agree with the experimentally measured values. For a wide range of quality factors QF_a for the JPEG compression, the cross-correlation decreases almost linearly with decreasing values of QF_a. For $QF_a = 5$, the cross-correlation with the embedded watermark is lost. Fig. 9 also includes curves for Gaussian and Laplacian models for the host-signal. These curves are similar to those in Fig. 4 in Section 4.2, except that the quality after quantization is indicated by the JPEG quality factor QF_a. The QF_a-values used in Fig. 9 can be translated to χ-values with help of Table 1. The theoretical results using the Gaussian or Laplacian model for the input sequence differ significantly from those found by the generalized Gaussian model or by actually measuring the cross-correlation. The results confirm the observation made in Section 4 that especially Gaussian host signals offer much larger robustness to quantization attacks than Laplacian signals or signals with even more peaky distributions.

Finally, we will briefly discuss the results for the sub-channel 22, as presented in Fig. 10. The watermark in this sub-channel is not robust at all due to the small variance of the host signal and the small subjective significance of these DCT-coefficients. The χ-values given in Table 1 show that JPEG applies to these coefficients strong quantization – even for high quality factors. As a result the cross-correlation decreases already significantly after JPEG compression with quality factors $QF_a > QF_e = 70$. The generalized Gaussian model agrees closely with the experimental results, but the Gaussian and Laplacian model do not. However, this result is hidden by the steep decrease of the cross-correlation in the range of a few quality factors.

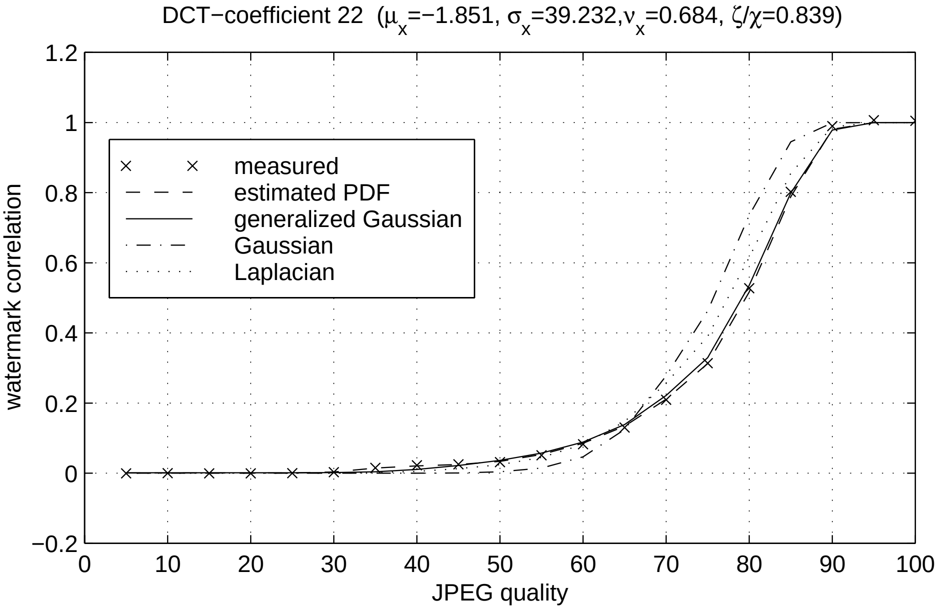

Fig. 10. Watermark correlation for sub-channel 22

5.5.3 *Standard Deviation of Watermark Correlation*

Besides the cross-correlation $\mathrm{E}\left\{\mathbf{yw}\right\}$, the standard deviation of the estimation of this cross-correlation has been measured. The correlation length, here equal to the sub-signal length $M = 1024$, and the availability of the original document have an important influence on the achievable standard deviation and thus on the detection error probability.

In Fig. 11 the experimental results and the results when using the generalized Gaussian model are shown for the sub-channel 1. In these plot the standard deviation of the experimentally measured correlation is indicated by a '+'-sign and a line above and underneath of each correlation value. The theoretical results are plotted similarly, however, with solid lines. Both hypotheses H_1 and H_0 are considered for detection with original. The plots show that the theoretical and experimental results agree approximately. Even better results can be achieved when using the estimated PDF instead of the generalized Gaussian model for the theoretical computation. Nevertheless, the generalized Gaussian model is sufficiently accurate. For large JPEG qualities Q_a, we see that the standard deviations for hypothesis H_1 and H_0 differ. This agrees with the observation made in Section 4 for fine quantization. The left plot in Fig. 11 indicates that the detection with original is relatively robust for large Q_a. However, blind detection is not possible for sub-channel 1 due to the small embedding strength. The right plot in Fig. 11 depicts for clarity only the results for hypothesis H_1 of the blind detection case. The standard deviation of

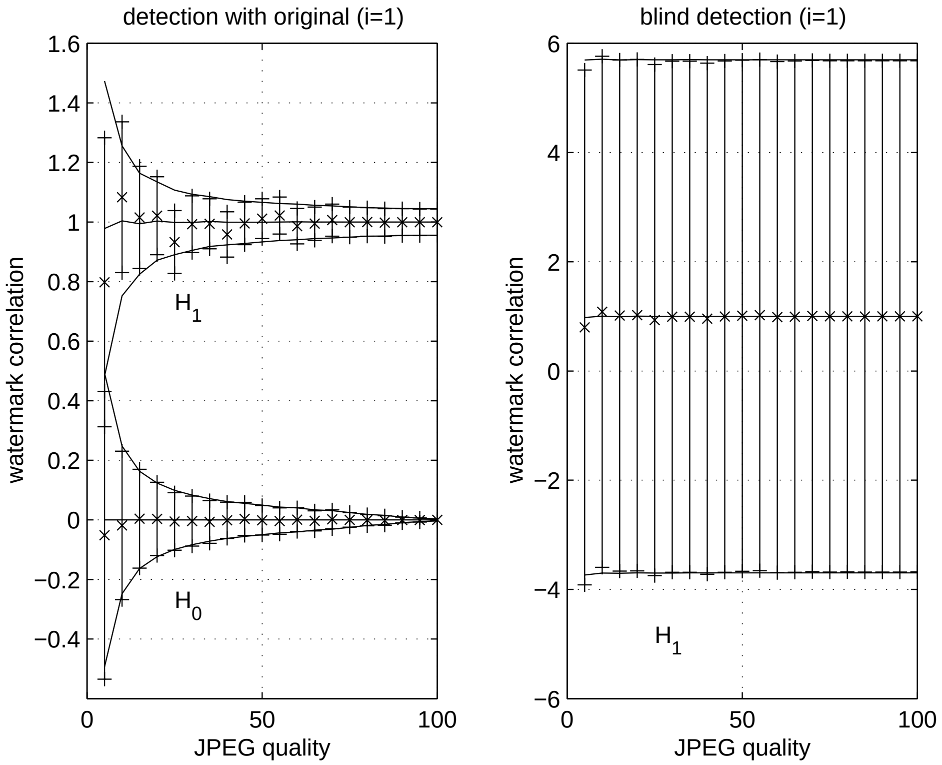

Fig. 11. Correlation mean and standard deviation for sub-channel 1 (DC value)

the correlation measurement is much larger than 1 and thus leads to high detection error probabilities.

In Fig. 12 the watermark detection for sub-channel 10 is shown in the same manner as for sub-channel 1 in Fig. 11. Both hypotheses H_1 and H_0 are depicted for detection with original (left plot) and for blind detection (right plot). Again, the theoretical results are derived using the generalized Gaussian model. We observe that the measured and predicted standard deviations are very similar. The right plot reveals that the standard deviation for hypothesis H_1 can even decrease for stronger quantization. This can be explained by the fact that strong quantization also reduces the interference from the original.

5.5.4 Sub-Channel Detection Error Probabilities

The presented results for the sub-channel watermark cross-correlation and its standard deviation enable the prediction of detection error probabilities, as described in Section 2. The experimental detection error rate of all 200 watermarks per sub-channel can be compared with the predicted values. However, only error probabilities down to 0.5% can be verified with this small number of experiments. Therefore, Fig. 13 and 14 show the measured error probabilities only for the plots with linear axis. The plots with logarithmic axis depict only error probabilities predicted assuming Gaussian distributed cross-correlation values. These error probabilities are

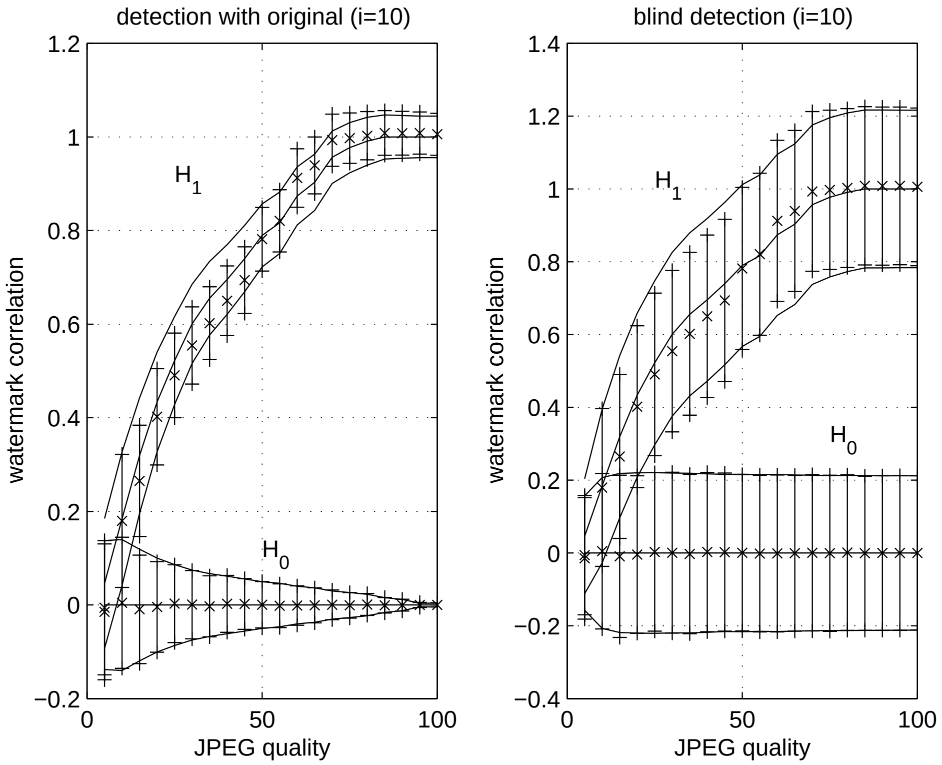

Fig. 12. Correlation mean and standard deviation for sub-channel 10

computed from the correlation mean $\mathrm{E}\left\{\mathbf{c}|\mathrm{H}_1\right\}$ and the variances Var$\left\{\mathbf{c}|\mathrm{H}_1\right\}$ and Var$\left\{\mathbf{c}|\mathrm{H}_0\right\}$.

Fig. 13 and 14 show the error probabilities for the watermark in the sub-channel 10. When detecting with the original, the error probability after a JPEG attack with $QF_a = 35$ (half of the embedding quality QF_e) is still lower than 10^{-5}. However, with a host signal of equal power but Gaussian distribution we could expect error probabilities around 10^{-9}, as indicated by the theoretically derived curve in Fig. 13. For blind detection, error probabilities down to 1% are measured. Note, that this is the result for one sub-channel. In practical applications, the cross-correlation in several sub-channels must be combined to achieve sufficiently low detection error probabilities (see e.g. [1]). Fig. 14 shows clearly that the measured detection error probabilities agree with those predicted using the generalized Gaussian model or by measuring the cross-correlation and its variance for all 200 simulations.

The error probabilities for sub-channel 1 and 22 are not plotted. Due to the small embedding strength for sub-channel 1, it is impossible to detect the watermark in this sub-channel with a blind detector. Using the original, it is possible to achieve low error probabilities down to quality factors $QF_a = 20$. The watermark in sub-channel 22 is not very robust. Detection with original has low error probabilities only after JPEG compression with $QF_a > QF_e = 70$.

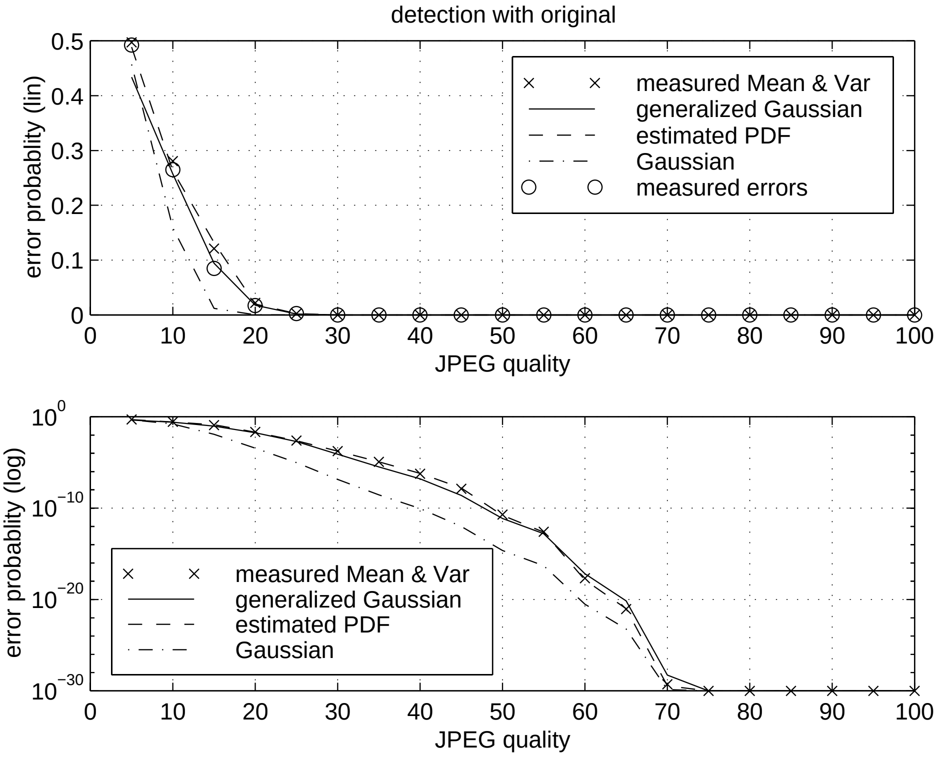

Fig. 13. Error probability for watermark detection with original (sub-channel 10)

5.5.5 *Summary*

The presented discussion could be extended to results for all different DCT coefficients, different embedding qualities and different watermark distributions. All experiments confirm that the theory presented in Section 3.2 and Section 4 allows the prediction of the robustness of correlation detection of an additive watermark after uniform scalar quantization attacks. Besides the absolute power of the host's sub-signals, the shape of the PDF for each of the sub-signals $\vec{x}_i$ has an important influence on the watermark detection robustness.

6 Conclusions

The major goal of this article is to analyze the effects of quantization on additive watermarking schemes. Particularly, uniform scalar quantization of watermarked documents with subsequent watermark detection using a correlation detector is investigated. A key factor for the analysis is the computation of statistical dependencies between the quantized watermarked document and the watermark itself. These dependencies have been derived in Section 3 by extending the theory of dithered quantizers. In Section 4 the performance of watermark detection using correlation is analyzed with help of the derived expressions. One important result is that the expected correlation decreases significantly for coarse quantization of non-uniformly

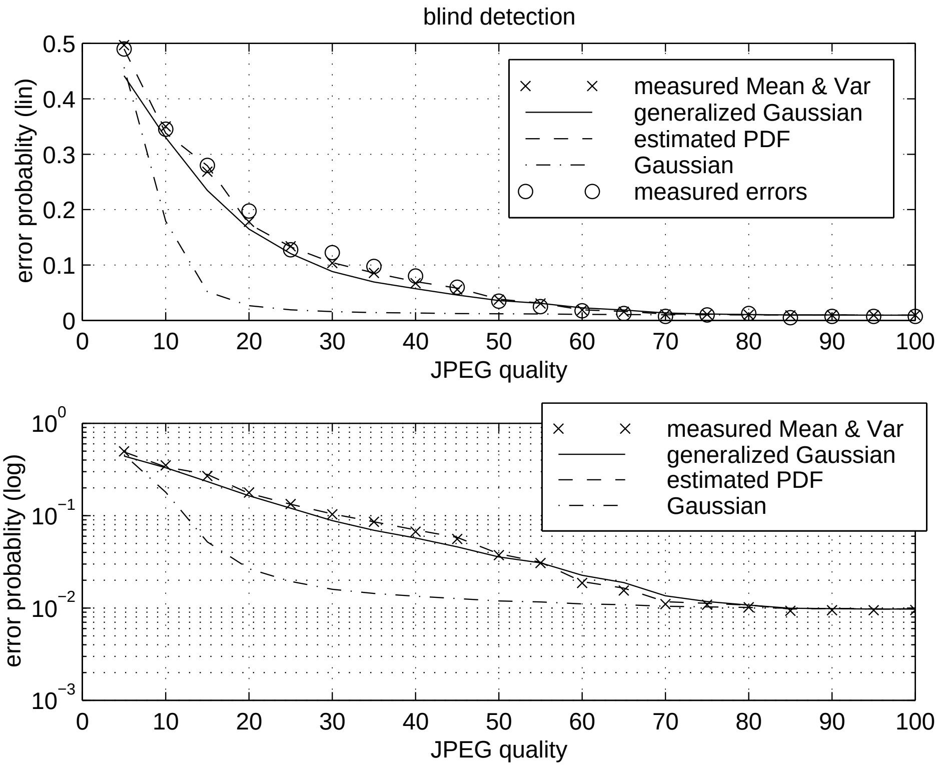

Fig. 14. Error probabilities for blind watermark detection (sub-channel 10)

distributed host signals. Thus, it is not possible to model the interfering noise at the watermark detector by AWGN. Although the investigated correlation depends on the distribution of the host document and the watermark, the distribution of the watermark does not play a significant role for common strengths of the watermark signal. However, the distribution of the host signal is very important. We showed that watermarks in Gaussian host signals are much more robust in the presence of quantization attacks than watermarks in Laplacian host signals. For the same host-document-to-noise ratio (DNR) after quantization, watermarks in Gaussian host signals can be detected with much lower error probability. The theoretically derived results are confirmed in Section 5 for an example image watermarking scheme under JPEG compression attacks. The experiments show that the parameters (coarseness of quantization, watermark embedding strength) considered for the theoretical analysis are likely to occur in practical environments. Further, the experiments show that the watermark detection error probability after JPEG compression can be predicted with sufficient accuracy using a generalized Gaussian model for the DCT coefficients. Thus, it is possible to compute the maximal robustness of a given watermark and given host image against JPEG compression.

Acknowledgements

The authors thank Jonathan K. Su for providing his software for the parameter estimation of the generalized Gaussian model and for many helpful discussions.

Appendix

A Summary of Formulas for the Statistical Signal Models Used

Gaussian, Laplacian, uniform and bipolar random variables are considered in this article as examples for the input signal $\mathbf{x}$ and the dither signal $\mathbf{d}$ (here equal to the watermark $\mathbf{w}$). To enable straightforward application of the presented theory, closed form expression of the considered PDFs and characteristic functions are summarized here.

Gaussian random variable:

$$p_{\mathbf{x}}(x) = \frac{1}{\sqrt{2\pi}\sigma_{\mathbf{x}}} \mathrm{e}^{-\frac{x^2}{2\sigma_{\mathbf{x}}^2}} \tag{A.1}$$

$$M_{\tilde{\mathbf{x}}}(ju) = \mathrm{e}^{-\frac{1}{2}u^2} \tag{A.2}$$

$$M_{\tilde{\mathbf{x}}}^{(1)}(ju) = ju\,\mathrm{e}^{-\frac{1}{2}u^2} \tag{A.3}$$

$$M_{\tilde{\mathbf{x}}}^{(2)}(ju) = (1-u^2)\,\mathrm{e}^{-\frac{1}{2}u^2} \tag{A.4}$$

$$M_{\tilde{\mathbf{x}}}^{(3)}(ju) = j(3u-u^3)\,\mathrm{e}^{-\frac{1}{2}u^2} \tag{A.5}$$

$$\mathrm{E}\left\{\mathbf{x}^4\right\} = 3\,\sigma_{\mathbf{x}}^4 \tag{A.6}$$

Laplacian random variable:

$$p_{\mathbf{x}}(x) = \frac{1}{\sqrt{2}\sigma_{\mathbf{x}}} \mathrm{e}^{-\frac{\sqrt{2}|x|}{\sigma_{\mathbf{x}}}} \tag{A.7}$$

$$M_{\tilde{\mathbf{x}}}(ju) = \frac{1}{1+\frac{1}{2}u^2} \tag{A.8}$$

uniform random variable:

$$p_{\mathbf{x}}(x) = \frac{1}{\sqrt{12}\sigma_{\mathbf{x}}} \mathrm{rect}\left(\frac{x}{\sqrt{12}\sigma_{\mathbf{x}}}\right) \tag{A.9}$$

$$
\begin{aligned}
M_{\tilde{\mathbf{x}}}\left(ju\right) &= \text{si}\left(\sqrt{3}u\right) = \frac{\sin\left(\sqrt{3}u\right)}{\sqrt{3}u} && \text{(A.10)}\\
M_{\tilde{\mathbf{x}}}^{(1)}\left(ju\right) &= j\frac{\sin\left(\sqrt{3}u\right)-\sqrt{3}u\cos\left(\sqrt{3}u\right)}{\sqrt{3}u^2} && \text{(A.11)}\\
M_{\tilde{\mathbf{x}}}^{(2)}\left(ju\right) &= \sqrt{3}\frac{12u^2\sin\left(\sqrt{3}u\right)-8\sqrt{3}u\cos\left(\sqrt{3}u\right)-8sin\left(\sqrt{3}u\right)}{12u^3} && \text{(A.12)}\\
M_{\tilde{\mathbf{x}}}^{(3)}\left(ju\right) &= j\frac{(6u-3u^3)\cos\left(\sqrt{3}u\right)+(3u^2-2)\sqrt{3}\sin\left(\sqrt{3}u\right)}{u^4} && \text{(A.13)}\\
\mathrm{E}\left\{\mathbf{x}^4\right\} &= 1.8\;\sigma_{\mathbf{x}}^4 && \text{(A.14)}
\end{aligned}
$$

bipolar random variable:

$$
\begin{aligned}
p_{\mathbf{x}}\left(x\right) &= 0.5\left(\delta\left(x-\sigma_{\mathbf{x}}\right)+\delta\left(x+\sigma_{\mathbf{x}}\right)\right) && \text{(A.15)}\\
M_{\tilde{\mathbf{x}}}\left(ju\right) &= M_{\tilde{\mathbf{x}}}^{(2)}\left(ju\right) = \cos(u) && \text{(A.16)}\\
M_{\tilde{\mathbf{x}}}^{(1)}\left(ju\right) &= M_{\tilde{\mathbf{x}}}^{(3)}\left(ju\right) = j\sin\left(u\right) && \text{(A.17)}\\
\mathrm{E}\left\{\mathbf{x}^4\right\} &= \sigma_{\mathbf{x}}^4 && \text{(A.18)}
\end{aligned}
$$

References

[1] J. J. Eggers and B. Girod, "Watermark Detection after Quantization Attacks," in *Proceedings Third International Workshop on Information Hiding*, Andreas Pfitzmann, Ed., Dresden, Germany, September/October 1999, vol. 1768, pp. 172–186, Lecture Notes in Computer Science, Springer.

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

[3] P. Moulin and J. A. O'Sullivan, "Information-Theoretic Analysis of Information Hiding," Preprint, September 1999.

[4] 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.

[5] 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.

[6] 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.

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

[8] 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.

[9] D. Kundur and D. Hatzinakos, “Digital watermarking using multiresolution wavelet decomposition,” in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing 1998 (ICASSP 98), Seattle, WA, USA*, May 1998, vol. 5, pp. 2969–2972.

[10] M. Kutter, F. Jordan, and F. Bossen, “Digital signature of color images using amplitude modulation,” in *Proc. of SPIE storage and retrieval for image and video databases*, San Jose, California, USA, February 1997, vol. 3022-5, pp. 518–552.

[11] M. Kutter and F. A. P. Petitcolas, “A fair benchmark for image watermarking systems,” in *Proc. of SPIE Vol. 3657: Security and Watermarking of Multimedia Contents*, San Jose, Ca, USA, January 1999, pp. 226–239.

[12] D. Tzovaras, N. Karagiannis, and M. G. Strintzis, “Robust Image Watermarking in the subband or discrete cosine transform domain,” in *Proceedings European Signal Processing Conference (EUSIPCO 98)*, Greece, September 1998, vol. 4, pp. 2285–2288.

[13] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, “Attacks on copyright marking systems,” in *Second International Workshop on Information Hiding, Proceedings published by Springer as Lecture Notes in Computer Science*, Portland, Oregon, USA, April 1998, vol. 1525, pp. 218–238.

[14] T. Kalker, J.-P. Linnartz, G. Depovere, and M. Maes, “On the reliability of detecting electronic watermarks in digital images,” in *Proceedings European Signal Processing Conference (EUSIPCO 98)*, Greece, September 1998, vol. 1, pp. 13–16.

[15] J.-P. Linnartz, T. Kalker, and J. Haitsma, “Detecting electronic watermarks in digital video,” in *Proceedings of the IEEE Intl. Conference on Speech and Signal Processing 1999 (ICASSP 99)*, Phoenix, USA, April 1999.

[16] L. Xie and G. R. Arce, “A Blind Wavelet Based Digital Signature for Image Authentication,” in *Proceedings European Signal Processing Conference (EUSIPCO 98)*, Greece, September 1998, vol. 1, pp. 21–24.

[17] J. R. Hernández and F. Pérez-González, “Statistical Analysis of Watermarking Schemes for Copyright Protection of Images,” *Proceedings of the IEEE, Special Issue on Identification and Protection of Multimedia Information*, vol. 87, no. 7, pp. 1142–1165, July 1999.

[18] F. Hartung, J. K. Su, and B. Girod, “Spread spectrum watermarking: Malicious attacks and counter-attacks,” in *Proc. of SPIE Vol. 3657: Security and Watermarking of Multimedia Contents*, San Jose, Ca, USA, January 1999, pp. 147–158.

[19] J. K. Su, F. Hartung, and B. Girod, “A channel model for a watermark attack,” in *Proc. of SPIE Vol. 3657: Security and Watermarking of Multimedia Contents*, San Jose, Ca, USA, January 1999, pp. 159–170.

[20] I. A. Glover and P. M. Grant, *Digital Communications*, Prentice Hall, London, New York, Toronto, Sydney, 1998.

[21] J. C. Hancock and P. A. Wintz, *Signal Detection Theory*, McGraw-Hill, Inc, New York, St. Louis, San Francisco, Toronto, London, Sydney, 1966.

[22] J. G. Proakis and D. G. Manolakis, *Digital signal processing, principles algorithms and applications*, Prentice Hall, 1996.

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

[24] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice Hall, 1984.

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

[26] G. Nicchiotti and E. Ottaviani, “Non-invertible statistical wavelet watermarking,” in *Proceedings European Signal Processing Conference (EUSIPCO 98)*, Greece, September 1998, vol. 4, pp. 2289–2292.

[27] Xiang-Gen Xia, C. G. Boncelet, and G. R. Arce, “A multiresolution watermark for digital images,” in *Proceedings of the IEEE Intl. Conference on Image Processing 1997 (ICIP 97)*, Santa Barbara, CA, USA, October 1997, vol. 1, pp. 548–551.

[28] G. K. Wallace, “The JPEG still picture compression standard.,” *Communications of the ACM*, vol. 34, no. 4, pp. 31–44, April 1991.

[29] J. K. Su and B. Girod, “Power-Spectrum Condition for Energy-Efficient Watermarking,” *IEEE Transactions on Multimedia*, 1999, Submitted.

[30] 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.

[31] K. A. Birney and T. R. Fischer, “On the modeling of DCT and subband image data for compresssion,” *IEEE Transactions On Image Processing*, vol. 4, no. 2, pp. 186–193, February 1995.

[32] F. Müller, “Distribution shape of two-dimensional DCT coefficients of natural images,” *Electronic Letters*, vol. 29, no. 22, pp. 1935–1936, October 1993.

[33] S. G. Mallat, “A theory for multiresolution signal decomposition: The wavelet representation,” *IEEE Transactions On Pattern And Machine Intelligence*, vol. 11, no. 7, pp. 674–693, July 1989.

List of Figures

List of Tables

List of Symbols

$\vec{x}$, $x[n]$, $\mathbf{x}$	host signal, sample, random variable
$\vec{w}$,$w[n]$, $\mathbf{w}$	watermark signal, sample, random variable
$\vec{s}$, $s[n]$, $\mathbf{s}$	public signal, sample, random variable
$\vec{r}$, $r[n]$, $\mathbf{r}$	received public signal, sample, random variable
$\vec{y}$, $y[n]$, $\mathbf{y}$	pre-processed received public signal, sample, random variable
$\vec{z}$, $z[n]$, $\mathbf{z}$	quantizer input signal, sample, random variable
$\vec{d}$, $d[n]$, $\mathbf{d}$	dither signal, sample, random variable
$\underline{\vec{z}}$, $\underline{z}[n]$, $\underline{\mathbf{z}}$	quantizer output signal, sample, random variable
$\underline{\underline{\vec{z}}}$, $\underline{\underline{z}}[n]$, $\underline{\underline{\mathbf{z}}}$	subtractive dithered quantizer output signal, sample, random variable
$\vec{e}$, $e[n]$, $\mathbf{e}$	subtractive quantization error / noise signal, sample, random variable
$\vec{\epsilon}$, $\epsilon[n]$, ϵ	non-subtractive quantization error signal, sample, random variable
Δ	quantizer step size
$\mu_{\mathbf{u}}$	mean of random variable $\mathbf{u}$
$\sigma_{\mathbf{u}}$	standard deviation of random variable $\mathbf{u}$
$\nu_{\mathbf{u}}$	shape factor of generalized Gaussian random variable $\mathbf{u}$
ζ	dither standard deviation normalized by quantizer step size
χ	host standard deviation normalized by quantizer step size
k	index of watermark or index for public signal with a certain watermark
i	sub-channel index
γ_x	scalar factor describing suppressed portion of host signal at the receiver
K	decision threshold for hypothesis test
p_{FP}	probability of false positive hypothesis decision
p_{FN}	probability of false negative hypothesis decision
c[n],$\mathbf{c}$	product sample $y[n]w[n]$, corresponding random variable
C,$\mathbf{C}$	correlation value, corresponding random variable
$p_{\mathbf{u}}\left(u\right)$	PDF of random variable $\mathbf{u}$
$p_{\tilde{\mathbf{u}}}\left(u\right)$	PDF of normalized random variable $\mathbf{u}$
$p_{\mathbf{u}}\left(u \middle\| v\right)$	PDF of random variable $\mathbf{u}$ dependent on v
$M_{\mathbf{v}}^{k}\left(ju\right)$	kth characteristic function of random variable $\mathbf{v}$ for u
$M_{\tilde{\mathbf{v}}}^{k}\left(ju\right)$	kth characteristic function of normalized random variable $\mathbf{v}$ for u
$\mathrm{E}\left\{\cdot\right\}$	expectation
$\mathrm{Var}\left\{\cdot\right\}$	variance
$\mathrm{STD}\left\{\cdot\right\}$	standard deviation