

# Performance Analysis of Information Hiding

Shuichi Shimizu

IBM Japan, Tokyo Research Laboratory  
1623-14 Shimotsuruma, Yamato-shi, Kanagawa 242-8502, Japan

shue@jp.ibm.com

## ABSTRACT

Information hiding in host data, or transparent digital watermarking, can be treated as an application of digital communications in which the hidden information is conveyed through a channel where the noise includes the host data and stems from other sources. The amount of information to be hidden is called the payload. At the detector, the hidden information (the watermark) should be retrieved with high confidence. We present a theoretical performance analysis of this information hiding problem in terms of payload, detection error rate, SNR, bandwidth of the watermarking channel, and channel coding for error correction. The detector is assumed to be a correlator, which is known to be optimal for Gaussian noise. However, our analysis does not require that the host data has a Gaussian distribution. Since our analysis does not depend on the synchronization between the watermark signal and the detector or on the maximum watermark power as constrained by preserving the fidelity, our result defines the theoretical performance limits. We present two decision rules designed to satisfy the given false alarm and code word error rate, based on energy detection and SNR estimation. We then apply two watermarking schemes, one with constant strength and the other with adaptive strength, in order to determine the watermarking design parameters by examining how the SNR is decreased against random and quantization noises.

**Keywords:** Digital watermarking, information hiding, payload, detection error, estimation, quantization noise

## 1. INTRODUCTION

*Information hiding*, the insertion of imperceptible digital watermark (invisible for video and inaudible for audio applications) into host media data, offers a way to carry secondary information such as an annotation which is tightly coupled to the primary host data. The maximum number of information bits which the watermark is able to carry is called the payload and is an important factor for digital watermarking. Payload limits in terms of channel capacity was discussed in previous work.<sup>1,2</sup> On the other hand, bit errors may occur when the information is retrieved during watermark detection, especially when one or more post processing steps such as lossy data compression have been applied, though the bit errors should be avoided or kept within a very low and tolerable rate. The robustness of watermarks is thus described in terms of a low rate of bit errors. The performance of error correction codes was discussed under Gaussian noise in previous work.<sup>3,4</sup> Another detection error is a false alarm when watermarks are detected from non-watermarked data, which is also an important issue for many applications such as copy protection.<sup>5</sup> The overlooking or miss of watermarks, when watermarks are not detected from watermarked data, is related to the bit error and false alarm rates, however, it is substantially dominated by the strengths of noise, which is not under our control because it depends on the post processing that is unpredictable.

This paper presents an analysis on watermarking performance in terms of the payload limit, minimally required watermark-signal-to-noise ratio (SNR), minimally required energy, probability of watermark detection error, bandwidth of the watermarking channel, and some channel coding parameters. Based on the analysis, the paper discusses two types of decision rules for detecting the existence of watermarks and for achieving a very low bit error rate. Also considered are the effects of random noise and quantization noise on how the noise decreases the watermark-signal-to-noise ratio, and which can be referred to for designing watermarking parameters such as payload and bandwidth.

In Section 2, we review the watermarking processes. In Section 3, we analyze the performance of watermarking by discussing detection error rate, and then we introduce the robustness margin for reliable detection. In Section 4, we give a practical analysis of decreasing SNR against random noise and quantization noise by introducing two watermarking schemes, the constant-strength and adaptive-strength schemes.

## 2. WATERMARK CREATION AND DETECTION

Digital watermarking may be classified as robust or fragile. In this paper, we shall be concerned only with robust watermarking, which refers to those methods where the inserted watermarks are still detectable after signal processing or deliberate attack. Most of the robust watermarking schemes<sup>6,7</sup> are based on some form of spread spectrum techniques, in which watermark creation and detection may be regarded as an application of digital communications. For these schemes, because a large amount of redundancy is required to overcome the effect of processing or attack on the watermark signal, the bandwidth  $W$  of the inserted watermark is much greater than the information rate  $R$ .

### 2.1. Watermark Creation

Watermark creation mainly involves the following three steps: (1) Channel coding generates a code word from the information bits. This code word can be viewed as a binary baseband signal. We denote by  $c_{ij}$  the  $j$ -th bit (0 or 1) in the  $i$ -th code word, where  $1 \leq i \leq 2^R$  and  $1 \leq j \leq W$ . We refer to the simple repetition of the bits to expand  $R$  to  $W$  as the uncoded case. Optionally, the  $W$  bits may be interleaved. (2) Spectrum spreading, or more specifically direct sequence spread spectrum (DSSS), multiplies a pseudo random number (PN) sequence of 0 and 1,  $\{b_{s+j}\}$ , where  $s$  is the starting offset in a very long PN sequence, by the baseband signal to generate another sequence,  $g_j = (2c_{ij} - 1)(2b_{s+j} - 1)$  for the  $i$ -th code word. The symbol unit of spreading is called a chip. The starting offset  $s$  should differ for each watermark, and if the offset is not known to the detector, it should be constant. In this case, one of the advantages of DSSS may be lost, as will be discussed in Section 3.1. (3) Modulation segments the host data into a number of blocks, applies a discrete Fourier transform (DFT) or other transforms to each of the blocks, and selects a frequency component to carry  $g_j$ . The  $j$ -th transmitted signal for the  $i$ -th code word is represented by

$$w_{ij} = (2c_{ij} - 1)(2b_{s+j} - 1)\mu_j, \quad (1)$$

where the positive number  $\mu_j$  is the strength of the watermark. The case when no transform is applied, or when a transform is applied to the host data without segmenting it into blocks is also treated as modulation in this paper.

Note that watermark insertion into the host data requires maintaining the transparency of the watermark so that it should not degrade the fidelity of the host data. The watermarking strength  $\mu_j$  can be seen as an individual adjustment in each chip for maintaining the fidelity. However, this is beyond the scope of this paper, and so we just introduce a peak signal-to-noise ratio (PSNR) or watermarking energy,  $\sum_{j=1}^W \mu_j^2$ , as the degradation parameter in the later sections.

### 2.2. Watermark Detection

Watermark detection is also composed of three steps, which are the inverses of those in the creation process. (1) Demodulating converts the observed waveform signal to baseband based on the carrier frequencies for the case of modulation with a carrier, or extracts the appropriate pixels or coefficients to make the baseband signal for the case of modulation without any carrier. For simplicity, we assume synchronization prior to the modulation step. Then, the  $j$ -th received signal is  $r_j = w_j + n_j$ , where the code word is unknown and  $n_j$  indicates the additive noise that has been introduced in the channel. (2) Despreading multiplies the baseband signal with the same PN sequence used in the creation stage. (3) If the signal has been interleaved in the creation stage, then it is deinterleaved first. Decoding determines the most likely embedded information bits by using a soft-decision

detection on the continuous baseband signal obtained through the above two steps. In this step, the correlation metric is calculated for the  $k$ -th code word as follows:

$$M_k = \sum_{j=1}^W (2c_{kj} - 1)(2b_{s+j} - 1)r_j, \quad (2)$$

and the one with the maximum value is selected.

### 3. BASIC PERFORMANCE ANALYSIS

In general, there is a trade-off between the payload and robustness. Specifically, the higher the payload, the less robust the watermark. In the following, we analyze watermarking performance in terms of the payload and robustness under a constant detection error rate.

#### 3.1. Detection Error Rate

For linear codes, the differences between the correlation metrics of any code words can be mapped into metrics of an all-zero code word ( $M_1$ ) and the corresponding code word ( $M_m$ ). Thus, we can assume the all-zero code word ( $c_{1j} = 0$ ) is transmitted for our analysis without losing any generality. The correlation metrics for the two are now calculated as follows:

$$M_1 = \sum_{j=1}^W \{\mu_j - (2b_{s+j} - 1)n_j\}, \quad (3)$$

$$M_m = \sum_{j=1}^W \{-(2c_{mj} - 1)\mu_j + (2c_{mj} - 1)(2b_{s+j} - 1)n_j\}, \quad (4)$$

and so the distance between any two of the code words can be represented by

$$D_m = M_1 - M_m = 2 \sum_{j=1}^W c_{mj}\mu_j - 2 \sum_{j=1}^W c_{mj}(2b_{s+j} - 1)n_j. \quad (5)$$

The first term can be considered as a random sampling without replacement from the finite population  $\{\mu_1, \dots, \mu_W\}$ , in which the elements are not random but fixed, because we are able to determine them. The mean and variance are calculated as follows:

$$\mathbb{E} \left[ 2 \sum_{j=1}^W c_{mj}\mu_j \right] = 2 \sum_{j=1}^W c_{mj} \mathbb{E}[\mu_j] = 2w_m \mu_s, \quad (6)$$

$$\text{Var} \left[ 2 \sum_{j=1}^W c_{mj}\mu_j \right] = 4w_m \frac{W - w_m}{W - 1} \sigma_s^2 \approx 4w_m \sigma_s^2, \quad (7)$$

where the weight  $w_m$  is the number of 1's in the code word ( $1 \ll w_m \ll W$ ), and  $\mu_s$  and  $\sigma_s$  are calculated by  $\mu_s = (1/W) \sum_{j=1}^W \mu_j$  and  $\sigma_s^2 = (1/W) \sum_{j=1}^W (\mu_j - \mu_s)^2$ , respectively. It should be noted that even if the elements are not fixed in the population set, but the mean  $\mu_s$  and variance  $\sigma_s$  are fixed instead, then the above equations still hold, and therefore we can set the mean and variance constant for the the watermarking signal strength set  $\{\mu_j\}$  in the following discussion.

The mean of the second term is zero because zero-mean noise is assumed ( $\mathbb{E}[n_j] = 0$ ), and the variance is calculated as  $4w_m \sigma_N^2$ , where  $\sigma_N^2$  is the variance of the noise, because the PN sequence has no correlation, i.e.,  $\mathbb{E}_s[(2b_{s+j} - 1)(2b_{s+k} - 1)] = 0$ , even though the noise correlates to other noise, e.g.,  $\mathbb{E}[n_j n_k] > 0$ . When the offset  $s$  is not known by the detector, then the detector and creator of the watermark have to share a fixed offset. In this case, the uncorrelation of the PN sequence no longer holds, i.e.,  $\mathbb{E}[(2b_{s+j} - 1)(2b_{s+k} - 1)] = \pm 1$ , but we expect that the fixed sequence will be able to cancel out the non-zero correlation of the noise within the summation on average and we still obtain a variance around  $4w_m \sigma_N^2$ .

Another justification for deriving the variance is that the code word error rate may be dominated by code words with smaller weights. For those code words, the chosen  $\mu_j$  (or  $n_j$ ) are sparse in the summation because the weight  $w_m$  is much less than the bandwidth  $W$ , or  $w_m \ll W$ . In addition, they may be interleaved, and so we can assume that they have no correlation with each other. In this justification, we still obtain the variance  $4w_m\sigma_s^2$  (or  $4w_m\sigma_n^2$ ).

The first and second terms are uncorrelated with each other because we assume the watermarking strength is not correlated with the noise, or  $E[\mu_j n_j] = 0$ , and so the mean and variance of the distance  $D_m$  are finally obtained by  $E[D_m] = 2w_m\mu_s$  and  $\text{Var}[D_m] \approx 4w_m(\sigma_s^2 + \sigma_n^2)$ , respectively. Now, the first and second terms can be seen as summations of many random variables ( $1 \ll w_m$ ), so they asymptotically follow the Gaussian distribution:

$$D_m \sim N(2w_m\mu_s, 4w_m(\sigma_s^2 + \sigma_n^2)). \quad (8)$$

Thus, the probability of code word error such that the  $m$ -th code word happens to be selected instead of the actually-transmitted all-zero code word is

$$P_2(w_m) = \text{Prob}[D_m < 0] = Q\left(\sqrt{\frac{\mu_s^2}{\sigma_s^2 + \sigma_n^2} w_m}\right), \quad (9)$$

where  $Q(x)$  is the tail of Gaussian density:

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt. \quad (10)$$

We denote by  $E_c$  the energy per chip,  $N_0$  the single-sided noise power spectral density,  $E_b$  the energy per bit, and  $R_c$  the code rate. Using this notation,  $E_c = R_c E_b$  and  $E_c/(N_0/2) = \mu_s^2/(\sigma_s^2 + \sigma_n^2)$ . Since  $S = E_b R$  and  $N = N_0 W$ , Equation (9) can be rewritten as

$$P_2(w_m) = Q\left(\sqrt{2\frac{S}{N}\frac{W}{R}R_c w_m}\right), \quad (11)$$

in which the SNR, processing gain  $W/R$ , and coding gain  $R_c w_m$  determine the probability of code word error. When the code is concatenated as an outer code with a binary repetition code, then the coding gain of the combined code is  $R_c w_m = R_c^o w_m^o$ , where  $R_c^o$  and  $w_m^o$  are the code rate and weight of the outer code. Note that it is not necessary that the additive noise follows a Gaussian distribution in this analysis. However, when the noise follows the Gaussian distribution, then it is known that the correlator produces the maximum SNR, and thus it is optimum.

### 3.1.1. Uncoded Case

Now, we discuss the uncoded DSSS watermarking channel as one of the practical schemes. The uncoded DSSS does not employ any code word and its robustness is supported by the DSSS only. In the simple repetition code (i.e., no code word), the minimum of  $w_m$  is  $W/R$ , and the code rate is  $R_c = R/W$ . The probability of decoding error on  $M = 2^R$  code words is bounded above by

$$P_M \leq \sum_{m=2}^M P_2(w_m) \simeq RQ\left(\sqrt{2\frac{S}{N}\frac{W}{R}}\right). \quad (12)$$

This can also be derived by noting that the probability of code word error is the inverse of non-error or  $P_M = 1 - (1 - P_b)^R \simeq RP_b$ , where the bit error rate is  $P_b = Q(\sqrt{2(S/N)(W/R)})$ , and the independence of the signal for each bit is assumed.

### 3.1.2. Convolutional-Coded Case

When the  $(n, k)$ -convolutional code is introduced in the watermark creation stage, and a soft-decision detection such as the Viterbi algorithm is used to extract the information bits, the upper bound of the equivalent bit error rate is

$$P_b \leq \frac{1}{k} \sum_{d=d_{\text{free}}}^{\infty} \beta_d P_2(d), \quad (13)$$

where  $k$  is the number of input bits for the convolutional encoder,  $d_{\text{free}}$  is the minimum free distance of the convolutional code, and the coefficients  $\{\beta_d\}$  are obtained from the transfer function of the code.<sup>8</sup> Thus, the probability of code word error can be bounded by

$$P_M = 1 - (1 - P_b)^R < \frac{R}{k} \sum_{d=d_{\text{free}}}^{\infty} \beta_d Q \left( \sqrt{2 \frac{S}{N} \frac{W}{R} R_c^\circ d} \right), \quad (14)$$

if the information bits are assumed to be independent.

### 3.2. Channel Capacity

Shannon channel capacity  $C$ <sup>8,9</sup> gives an upper bound of bit rate for a given SNR with additive white Gaussian noise (AWGN) for any low bit error rate, as

$$R \leq C = W \log_2 \left( 1 + \frac{S}{N} \right). \quad (15)$$

The unit of channel capacity, bits per sec, should be translated to bits per bandwidth  $W$ . Equation (15) is for the continuous input and output to the channel. However, in the case of watermarks, the input to the channel is discrete ( $-A$  or  $+A$ ) while the output is continuous ( $\{-\infty, \infty\}$ ), and thus the channel capacity for the discrete input should be considered when representing a watermarking channel. For BPSK (binary phase-shift keying) with input  $= \pm A$  and AWGN, the channel capacity is given by

$$C = \frac{1}{2} \int_{-\infty}^{\infty} \left\{ p(y|A) \log_2 \frac{p(y|A)}{p(y)} + p(y|-A) \log_2 \frac{p(y|-A)}{p(y)} \right\} dy. \quad (16)$$

This can be calculated explicitly by using:

$$p(y|\pm A) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y\pm A)^2/2\sigma^2}, \quad (17)$$

$$p(y) = \frac{1}{2} \{p(y|A) + p(y|-A)\}, \quad (18)$$

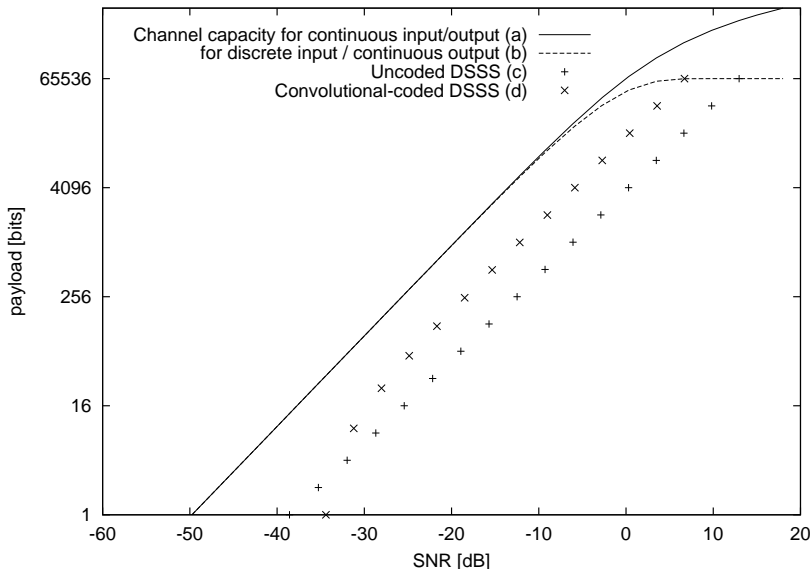
where  $\sigma^2$  is the variance of the AWGN. Note that  $C$  here represents bits per channel use, and thus it should be multiplied by the bandwidth  $W$  before it is compared to Equation (15). The SNR here is  $S/N = A^2/2\sigma^2$ , and thus the channel capacity  $C$  is a function of the SNR as well as in Equation (15).

### 3.3. Quantitative Trade-Offs

When the maximum error rate  $P_M$ , bandwidth  $W$ , and the channel coding parameters  $k$ ,  $\beta_d$ ,  $R_c^\circ$  are given, then the maximum payload  $R$  and minimally required SNR (or minimum SNR)  $S/N$  are determined from Equation (12) or (14), which are plotted in Fig. 1 with the inequality sign replaced by an equals sign.\* Also included in Fig. 1 is the Shannon channel capacity as a reference for the payload upper limits. For example, when the convolutional-coded DSSS is employed, then a 256-bit payload is obtained with  $-18.5$  dB as the minimally required SNR. When the error rate and/or the bandwidth are allowed to be greater, then the maximum payload

\*In (14),  $R$  is the payload plus the constraint length  $K$ , and the summation is calculated until  $d = 64$  because the terms for  $d > 64$  are small enough with comparison to the given error rate for drawing Fig. 1.

can be greater and/or the minimally required SNR can be smaller. The SNR is closely related to the robustness of watermarks. Just after watermark is embedded, the initial SNR is the ratio of the watermark signal to the power of the original image or audio data. Thus, the difference between the minimum SNR and the initial SNR is a robustness margin for the attenuation of watermark signals or the additive noise from unknown post processing such as lossy compression.



**Figure 1.** Trade-offs between the payload and minimum SNR for (a) Shannon channel capacity with continuous input and output, (b) channel capacity with discrete input, (c) the uncoded DSSS, and (d) the convolutional-coded DSSS with rate  $R_c^o = k/n = 1/2$ , constraint length  $K = 7$ , minimum free distance  $d_{\text{free}} = 10$ , coefficients  $\beta_{10} = 36$ ,  $\beta_{11} = 0$ ,  $\beta_{12} = 211, \dots$ , bandwidth  $W = 65, 536$ , and error rate  $P_M = 10^{-5}$ .

Note that one of the well-known watermarking schemes<sup>6</sup> employs  $M$ -ary orthogonal signals with  $M = 1,000$  (approximately a 10-bit payload) and bandwidth  $W = 1,000$ . Assuming an AWGN, then the minimum SNR is about  $-15.1$  dB for the  $10^{-5}$  error rate.<sup>†</sup> By shifting Fig. 1 by about  $+18.2$  dB (i.e.,  $65,536/1,000$ ), it is found that the minimum SNR for the uncoded DSSS is  $-9.5$  dB for an error rate of  $10^{-5}$ . Thus, the  $M$ -ary orthogonal channel coding contributes about 5.6 dB of gain when using this scheme. However, the use of larger  $M$  is computationally impractical.

### 3.4. Decision Rules

We introduce two decision rules for determining whether we should accept or reject watermarks under the given detection error rates. One is for a low code word error rate, and the other is for a low false alarm rate.

#### 3.4.1. Code Word Error

In practice, the SNR is unknown to the detector, and so it is necessary to estimate the SNR from the received signals or  $z_{lj} = (2c_{lj} - 1)(2b_{s+j} - 1)r_j$ , where  $l$  is the index of the selected code word, i.e.,  $l = \arg \max_k M_k$ . The sample mean  $\hat{\mu}_{z_l}$  and variance  $\hat{\sigma}_{z_l}^2$  are calculated as

$$\hat{\mu}_{z_l} = \frac{1}{W} \sum_{j=1}^W z_{lj}, \quad (19)$$

$${}^\dagger P_M = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \{1 - (1 - Q(y))^{M-1}\} \exp\left\{-\frac{1}{2} \left(y - \sqrt{2(S/N)W}\right)^2\right\} dy$$

$$\hat{\sigma}_{z_l}^2 = \frac{1}{W} \sum_{j=1}^W (z_{lj} - \hat{\mu}_{z_l})^2, \quad (20)$$

which are the minimum-variance unbiased estimators, and thus the SNR can be estimated by  $\hat{\mu}_{z_l}^2 / (2\hat{\sigma}_{z_l}^2)$ . The decision rule to accept the output from the decoder is then

$$\delta = \begin{cases} 1 & \text{if } \hat{\gamma}_l \triangleq \hat{\mu}_{z_l}^2 / (2\hat{\sigma}_{z_l}^2) \geq T, \\ 0 & \text{otherwise,} \end{cases} \quad (21)$$

where the threshold  $T$  corresponds to the minimum SNR, which is derived from the acceptable code word error rate and the other watermarking parameters such as the payload, as shown in Fig. 1.

### 3.4.2. False Alarm

Equation (21) is designed to be optimum for the binary hypothesis testing of acceptance of a code word, but it should not be applied as a decision rule for binary hypothesis testing that tests the existence or absence of a watermark, because it is not optimum for that purpose. The false alarm rate might be much higher than the code word error rate if the above decision rule is applied, because the maximization of  $2^R$  candidates increases it by about  $2^R$  times, which may not be desirable, especially for larger payloads.

Energy detection<sup>8</sup> is much better for avoiding a high false alarm rate. Suppose that a channel coding (outer code) is followed by repetition coding (inner code), and their code rates are  $R_c^o$  and  $R_c^i$ , respectively. Usually we have  $R/W < R_c^o, R_c^i < 1$ . However, if no channel coding is applied, then  $R_c^o = 1$  and  $R_c^i = R/W$ . Now, assuming that the received signal which had not been watermarked at the transmitter and which has been despread and deinterleaved at the receiver,  $z_j = (2b_{s+j} - 1)r_j$ , is individually and independently distributed (i.i.d.) with zero mean, then the signal summation for each bit in the outer code word asymptotically follows the Gaussian distribution, or  $\sum_{j \in \{i\text{-th bit}\}} z_j \sim N(0, \sigma^2/R_c^i)$ , where the variance is approximated by  $\sigma^2 \approx (1/W) \sum_{j=1}^W z_j^2$ . Let  $x_i$  denote the signal summation normalized by  $\sigma/\sqrt{R_c^i}$  for  $i$ -th bit, so  $x_i$  is the random variable which asymptotically follows the normalized Gaussian distribution, or  $x_i \sim N(0, 1^2)$ . The summation of  $x_i^2$  then follows the Chi-squared distribution with degree of freedom  $\nu = R/R_c^o$ , or  $\sum_{i=1}^{R/R_c^o} x_i^2 \sim \chi_\nu^2$ , where the mean and variance are  $R/R_c^o$  and  $2R/R_c^o$ , respectively. When the number of degrees of freedom  $\nu$  is large enough, then the false alarm rate is calculated by

$$P_{\text{FA}} = \text{Prob} \left[ \sum_{i=1}^{R/R_c^o} x_i^2 \geq T_{\text{FA}} \right] \simeq Q \left( \frac{T_{\text{FA}} - R/R_c^o}{\sqrt{2R/R_c^o}} \right). \quad (22)$$

Conversely, when the desired false alarm rate  $P_{\text{FA}}$  is given, then we can calculate the corresponding threshold  $T_{\text{FA}}$  by using Equation (22).

The square summation of  $x_i$  can be seen as the energy, and the decision rule for the energy detector is

$$\delta = \begin{cases} 1 & \text{if } \sum_{i=1}^{R/R_c^o} x_i^2 \geq T_{\text{FA}}, \\ 0 & \text{otherwise.} \end{cases} \quad (23)$$

Note that the energy detector is the estimator-correlator for a random Gaussian signal in the Gaussian noise and is well-known to be optimum from the viewpoint of a Neyman-Pearson detector.<sup>10</sup> The watermark signal is not actually random, however, and when the information rate  $R$  is large enough, then it can be approximated as a random signal, and thus the energy detector is near optimum for large  $R$ .

When applying both decision rules, the decision rule (23) for the existence test, which is based on the energy detector, should be applied first. When the received signal doesn't pass the existence test, then it is rejected as "no watermark found". Otherwise, it is forwarded to the second test, the decision rule (21) for satisfying the very low code word error rate, which is based on the estimates of SNR. When it passes the second test, then it is accepted and interpreted as information bits. Otherwise, it is rejected as "watermark found, but unable to decode". The probability of missing of the watermark is thus dominated by the energy and SNR of the watermark signal.

## 4. ANALYSIS OF DECREASING SNR

In the previous sections, we mentioned that the SNR directly determines the code word error rate and that an estimate of the SNR is needed for deciding on the acceptance of the received signal as a watermark. In general, the SNR will decrease when one or more post signal processing steps have been applied to the transmitted signal. It is important to know how the SNR decreases in various situations so that we can determine appropriate watermarking parameters to make the watermarks robust. Below, we discuss two types of noise, random noise and quantization noise in noisy channels.

### 4.1. Random Noise

Supposing that zero-mean random noise with variance  $\sigma_R^2$  is added in the channel, increasing the noise power from its initial  $\sigma_S^2 + \sigma_N^2$  to  $\sigma_S^2 + \sigma_N^2 + \sigma_R^2$ . When we have an  $m$ -dB robustness margin, then the allowable power of the random noise,  $\sigma_R^2$ , is described as

$$\frac{\sigma_R^2}{\sigma_S^2 + \sigma_N^2} \leq 10^{\frac{m}{10}} - 1. \quad (24)$$

For example, if the robustness margin is  $m = 6$  [dB], then random noise is allowed up to about three times the initial noise level.

### 4.2. Quantization Noise

Quantization discretizes the amplitude while sampling discretizes the time, and it is, for example, performed in a discrete cosine transform (DCT) domain for JPEG and MPEG.<sup>11</sup> According to quantization theory,<sup>12</sup> the probability density function (PDF) of the continuous input signal is recoverable from the quantized output signal — in particular the mean of the PDF is preserved — if the input PDF is band-limited in frequency and the quantization step size  $q$  is small enough in comparison to the highest frequency  $\psi$  of the PDF or  $q < 2\pi/\psi$ . This means that the quantization does not affect the detection of watermarks much as long as the quantization step is small. However, because higher compression may be required in many situations for JPEG and MPEG, we should focus on larger quantization steps, where the above assumption does not hold.

Given a random variable  $Y$  with a continuous PDF  $P_Y(y)$ , a uniform quantizer with step size  $q$  produces a discrete PDF  $P_Z(z)$ . The first and second moments of the quantizer output are

$$E[Z] = \sum_{i=-\infty}^{\infty} iq \int_{(i-\frac{1}{2})q}^{(i+\frac{1}{2})q} P_Y(y) dy, \quad \text{and} \quad (25)$$

$$E[Z^2] = \sum_{i=-\infty}^{\infty} (iq)^2 \int_{(i-\frac{1}{2})q}^{(i+\frac{1}{2})q} P_Y(y) dy. \quad (26)$$

The mean and variance are then  $\mu_Z = E[Z]$  and  $\sigma_Z^2 = E[Z^2] - E^2[Z]$ , respectively, and the SNR is represented by  $\mu_Z^2/(2\sigma_Z^2)$ . If the mean decreases or  $E[Z] < E[Y]$ , then the quantization noise is no longer of zero mean but it may attenuate the watermarking power. Even in this case, Equation (12) still holds for representing the relation of the code word error rate and the SNR because it can be treated as a decrease of the mean watermarking-strength  $\mu_S$ .

We shall assume that the original host signal follows a generalized Gaussian PDF:

$$P_X(x) = \frac{c_1(\beta)}{\sigma} \exp\left(-c_2(\beta) \left|\frac{x}{\sigma}\right|^{\frac{1}{\beta}}\right), \quad (27)$$

where

$$c_1(\beta) = \frac{\Gamma(\frac{1}{2})(3\beta)}{2\beta\Gamma(\frac{3}{2})(\beta)}, \quad c_2(\beta) = \left[\frac{\Gamma(3\beta)}{\Gamma(\beta)}\right]^{\frac{1}{2\beta}} \quad (28)$$

and  $\Gamma(x)$  is the Gamma function. Note that  $\beta = 1/2$  represents the Gaussian PDF and  $\beta = 1$  represents the Laplacian PDF.

### 4.2.1. Constant- and Adaptive-Strength Watermarking Schemes

We introduce two watermarking schemes, the constant-strength watermarking (CSW) scheme and the adaptive-strength watermarking (ASW) scheme

$$Y = X + a\sigma, \quad (29)$$

$$Y = X + a|X|, \quad (30)$$

respectively, where the strength of watermark signals is constant for the former scheme ( $\mu_s = a\sigma$  and  $\sigma_s^2 = 0$ ), and is proportional to the strength of the host signals for the latter scheme. It should be noted that, in the ASW scheme, for a certain  $\beta$ , the mean and variance of the watermarking strength,  $\mu_s$  and  $\sigma_s^2$ , are constant because the PDF is fixed, and so the necessary conditions discussed in Section 3.1 hold. It should be also noted that the PSNR or watermarking energy is the same for the two schemes because  $\sum (|X|)^2 = \sum X^2 = \sum \sigma^2$ .

The PDF of the random variable  $Y$  for the CSW is

$$P_Y(y) = P_X(y - a\sigma), \quad (31)$$

and for the ASW the PDF is

$$P_Y(y) = \begin{cases} P_X\left(\frac{y}{1+a}\right) / (1+a) & y \geq 0, \\ P_X\left(\frac{y}{1-a}\right) / (1-a) & y < 0, \end{cases}, \quad (32)$$

where it is assumed that  $0 < a < 1$ .

### 4.2.2. SNR vs. Quantization Step Size

These equations allow us to compute the watermarking SNR vs. the normalized quantization step size. The solid line (a) in Fig. 2 is based on this analysis, showing the SNR decreasing with increasing quantization step size for the ‘‘Lena’’ image<sup>13</sup> with the CSW scheme by substituting Equation (31) into Equation (25) and (26). In the figure, the scaling factor<sup>‡</sup> is set to  $a = 0.5$ , and the horizontal axis is normalized to  $q/\sigma$ . This shows that the SNR rapidly decreases when the quantization step size exceeds twice the watermark strength or  $q/\sigma > 2a$ .

The points marked with ‘‘+’’ are from measurements. For the experimental measurements, the watermark was created on 16 coefficients in the middle range of each  $8 \times 8$ -DCT block. In general, the variance is greater for a lower frequency than for a higher frequency in natural images, and so the 16 coefficients are separately manipulated to represent the watermark so that the amount of manipulation would be proportional to the standard deviation  $\sigma_i$  of the  $i$ -th coefficient or  $Y_i = X_i + a\sigma_i$  ( $i = 1, \dots, 16$ ) for the CSW scheme. So that the watermark signal would be maximized in the detection stage, the quantizer outputs  $\{Z_i\}$  are normalized prior to summing them up by dividing them by the sample standard deviation  $(\hat{\sigma}_z)_i$ , where  $(x)_i$  indicates the statistic for the  $i$ -th coefficient. Thus, the SNR is estimated by

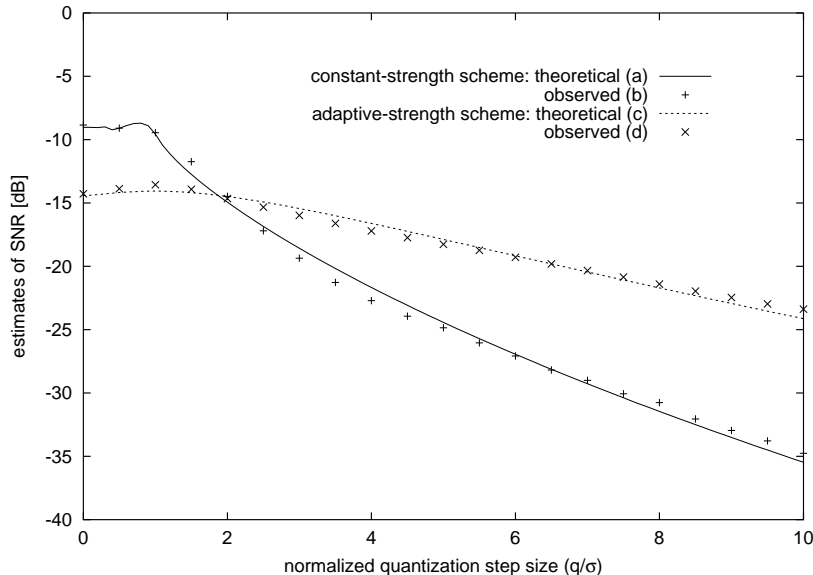
$$\hat{\gamma} = \frac{1}{2} \left[ \frac{1}{n} \sum_{i=1}^n \left( \frac{\hat{\mu}_z}{(\hat{\sigma}_z)_i} \right) \right]^2, \quad (33)$$

where  $n$  is the number of coefficients manipulated in each DCT block (e.g., 16).

The dotted line (c) shows the analysis for the ASW scheme. It is obtained by substituting Equation (32) into Equation (25) and (26) in the same way as for the CSW scheme. The points marked with ‘‘x’’ are from measurements. The deviation caused by various assignments of PN sequences and code words is negligible because of the high bandwidth.

Figures 1 and 2 show that the 256-bit payload in the bandwidth  $W = 65, 536$  can be embedded as long as the quantization step size is less than about 6 times the standard deviation of the host signal for the ASW scheme

<sup>‡</sup>The scaling factor  $a = 0.5$  makes the PSNR 42 [dB] with 4 [dB] standard deviation for the 100 sample natural images used in the experimental measurements. In the PSNR, the signal is the (peak) power of the image, and the noise is the power of the watermark.



**Figure 2.** Theoretical and observed estimates of SNR on the “Lena” image for the constant-strength and adaptive-strength schemes, in which a generalized Gaussian PDF is assumed, and  $\beta = 1.77$  is fitted to the image. ( $a = 0.5$ )

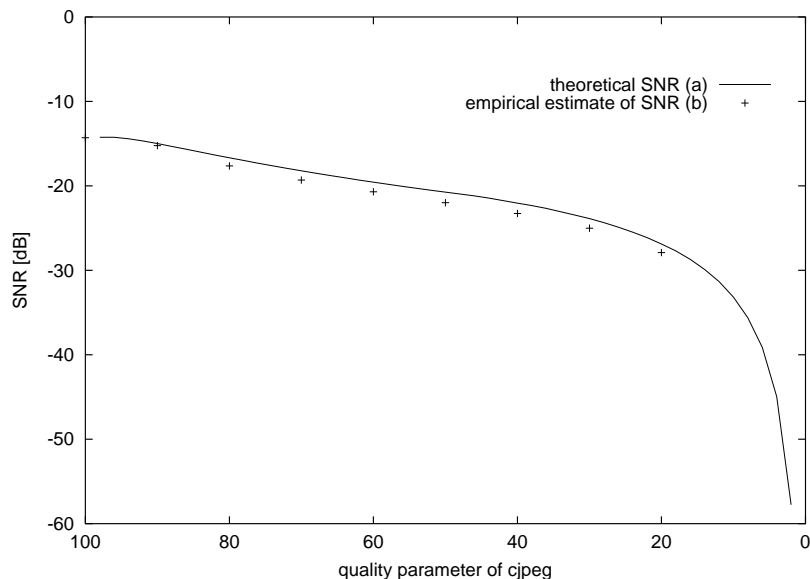
with the watermark strength  $a = 0.5$ , because the SNR is able to decrease down to about  $-18.5$  dB, as shown by Fig. 1. If more robustness is required, then we can change one or more of the watermarking parameters, for example, reduce the payload or increase the watermarking strength. In our sample of 100 natural images, the difference between analysis and measurement has a mean of  $-0.03$  [dB] with  $1.60$  [dB] as standard deviation, which indicates that the theoretical estimates on average fit the observed estimates.

It is also found that the ASW scheme is more robust than the CSW scheme in the ranges which we should focus on for practical situations, for example, for JPEG compression. On the other hand, for random noise, the CSW scheme is more robust than the ASW scheme for any range of noise strength, because the CSW scheme has no variance in the watermarking strength ( $\sigma_s^2 = 0$ ), which would act as noise, and it has less noise in every range. However, robustness against quantization-based lossy compression is usually more important than against a large amount of random noise, so the ASW scheme seems preferable for the practical situations.

In the case of “cjpeg”, which is a practical JPEG compression reference software,<sup>14</sup> the quantization step size is not the same for each coefficient but rather weighted so that larger quantization step sizes will be assigned for the lower frequencies. The set of weight assignments is called a quantization table, and each assignment is increased or decreased via a single parameter called the quality factor. Fig. 3 shows how the SNR decreases relative to the quality parameter in theoretical (solid line) and observed (‘+’ marks) plots. The estimate of the SNR is calculated by using Equation (33) for the theoretical line with different quantization steps for each coefficient. This shows that the watermark with a  $-18.5$  dB minimum SNR will survive up to the quality parameter of about 68 (or 75 from the empirical data), which will reduce the image size to about  $1/26$  (or  $1/23$ ) for the “Lena” ppm image.

## 5. CONCLUSION

We have analyzed the performance of information hiding in terms of the payload limit, the minimum SNR, the probability of detection error, the bandwidth of the watermarking channel, and the parameters of the channel coding, where the robustness of watermarks corresponds to realizing a very low detection error rate. Based on the analysis, we have shown two decision rules for accepting watermarks utilizing the energy detection and the estimate of the SNR so that they can provide very low rates of false alarms and bit errors. We have illustrated



**Figure 3.** Decreasement of theoretical SNR and empirical estimate of SNR against “quality” parameter of “cjpeg” on a “Lena” image, in which a generalized Gaussian PDF is assumed, and  $\beta = 1.77$  is fitted to the image, and the scaling parameter is set to  $a = 0.5$ .

how the analytical results can be applied in practice in terms of selecting parameters for watermarking, such as the payload and watermarking strength, by analyzing how the SNR will decrease under random noise and quantization noise when the distribution parameters of the host data are given.

### Acknowledgments

The author would like to thank Professor Bede Liu and other colleagues at Princeton University for helpful discussions. The author is also grateful to the colleagues in Tokyo Research Laboratory of IBM Japan for a lot of discussions.

### REFERENCES

1. M. Barni, F. Bartolini, A. de Rosa, and A. Piva, “Capacity of the watermark channel: How many bits can be hidden within a digital image?,” in *Proc. SPIE*, **3657**, pp. 437–448, (San Jose, CA), Jan. 1999.
2. P. Moulin, M. Mihcak, and G.-I. Lin, “An information-theoretic model for image watermarking and data hiding,” in *Proc. ICIP’00*, Oct. 2000.
3. F. P. Gonzalez, J. R. Hernandez, and F. Balado, “Approaching the capacity limit in image watermarking: a perspective on coding techniques for data hiding applications,” *Signal Processing* **81**, pp. 1215–1238, 2001.
4. S. Baudry, J. F. Delaigle, B. Sankur, B. Macq, and H. Maitre, “Analysis of error correction strategies for typical communication channels in watermarking,” *Signal Processing* **81**, pp. 1239–1250, 2001.
5. Secure Digital Music Initiative (SDMI). <http://www.sdmi.org>.
6. I. Cox, J. Kilian, T. Leighton, and T. Shanon, “Secure spread spectrum watermarking for multimedia,” *IEEE Trans. on Image Processing* **6**, pp. 1673–1687, Dec. 1997.
7. M. Barni, F. Bartolini, and A. Piva, “A dct-domain system for robust image watermarking,” *Signal Processing* **66**, pp. 357–372, May 1998.
8. J. Proakis, *Digital Communications, 3rd ed.*, New York: McGrawHill, 1995.
9. C. E. Shannon, “Communication in the presence of noise,” in *Proc. of the IRE*, **37**, pp. 10–21, 1949.

10. S. M. Kay, *Fundamentals of Statistical Signal Processing, Detection Theory, Volume II*, New Jersey: Prentice Hall, 1998.
11. G. K. Wallace, "The jpeg still picture compression standard," *Communications of the ACM* **34**(4), pp. 30–44, 1991.
12. B. Widrow, I. Kollar, and M. Liu, "Statistical theory of quantization," *IEEE Trans. on Instrumentation and Measurement* **45**(6), pp. 353–361, 1995.
13. USC-SIPI Image Database. <http://sipi.usc.edu/services/database/Database.html>.
14. Independent JPEG Group. <http://www.ijg.org>.