Improved QIM Strategies for Gaussian Watermarking[*]

Pierre Moulin and Ying Wang

University of Illinois, Beckman Inst., Coord. Sci. Lab & ECE Dept.,
405 N. Mathews Ave., Urbana, IL 61801
moulin@ifp.uiuc.edu

Abstract. This paper revisits the problem of watermarking a Gaussian host, where the embedder and attacker are subject to mean-squared distortion constraints. The worst (nonadditive) attack and unconstrained capacity have been identified in previous work. Here we constrain the encoding function to lie in a given family of encoding functions — such as spread-spectrum or fixed-dimensional Quantization Index Modulation (QIM), with or without time-sharing, with or without external dithering. This gives rise to the notion of constrained capacity. Several such families are considered in this paper, and the one that is best under the worst attack is identified for each admissible value of the watermark-to-noise ratio (WNR) and the noise-to-host ratio (NHR). With suitable improvements, even scalar QIM can outperform any (improved) spread-spectrum scheme, for any value of WNR and NHR. The remaining gap to unconstrained capacity can be bridged using higher-dimensional lattice QIM.

1 Introduction

Quantization-index modulation (QIM) methods, introduced by Chen and Wornell [1], possess attractive practical and theoretical properties for watermarking. On the practical side, they are easy to implement when scalar quantizers or some low-dimensional lattice quantizers are used. On the theoretical side, the *dithered* version of QIM (using an external, uniformly distributed dither vector shared by encoder and decoder) studied by Eggers *et al* [2] and by Erez and Zamir [3,4,5] is mathematically more tractable than the original (nondithered) QIM. It follows from Erez and Zamir's work [4,5] that there exist capacity-approaching lattice QIM coding and decoding schemes for data hiding under additive white Gaussian noise (AWGN) attacks, under some mild technical conditions on the host (*aka* interference) signal statistics. A remarkable byproduct of Erez and Zamir's analysis is that lattice QIM decoding causes no capacity loss vis-a-vis optimal maximum-likelihood (ML) decoding. In other words, the lattice QIM decoder is not penalized (in terms of achievable rates) by not knowing, or ignoring, host signal statistics.

[*] This work was supported by NSF under grant CCR 03-25924.

M. Barni et al. (Eds.): IWDW 2005, LNCS 3710, pp. 372–386, 2005.

Recent research in watermarking [6,7,8] has however raised the concern that for “weak host signals”, dithered scalar QIM performs somewhat poorly. In particular, [7] showed that dithered scalar QIM can be outperformed by spread-spectrum modulation (SSM) [9] methods in this scenario, and [8] studied possible improvements. The results in [6,7,8] were however restricted to the case of AWGN attacks. This is a rather restrictive assumption because if the host signal is weak, compression-type attacks are much more effective than additive-noise attacks. More precise formulations of this statement appear in [10,11,12].

This motivated us to revisit the data-hiding game under squared-error distortion constraints studied in [12] (and applied to image watermarking in [13,14]), in which the host signal is Gaussian, and all distortions are measured in an expected sense, *with respect to the host*. In this setup, the worst attack (in the sense of achieving unconstrained capacity) is the *Gaussian test channel* from rate-distortion theory [15], and the AWGN attack may be severely suboptimal. We then ask what is the performance of lattice QIM schemes of arbitrary dimensions, and whether substantial improvements are possible:

- either by exploiting the host signal statistics in the design of the QIM decoder,
- or by using an improved version of the QIM encoder,
- or both.

These lattice QIM schemes are compared with spread-spectrum schemes using linear precancellation of the host signal [16,17,18].

In addition to the standard QIM and SSM, several new or uncommon acronyms are used in this paper. They are summarized in Table 1 for convenience.

Table 1. List of acronyms used in this paper

Acronym	Full name	Equation
WNR	watermark to noise ratio	(7)
NHR	noise to host ratio	(8)
WNR_{eff}	effective WNR	(17)
aSSM	attenuated SSM	(19)
ISS	improved SSM [18]	Sec. 4
aQIM	attenuated QIM	(28)

2 Background: Mutual-Information Game

This section reviews some results from [11,12]. We use uppercase letters for random variables, lowercase for their individual realizations, and boldface for vectors. The symbol $\mathbb{E}$ denotes mathematical expectation. The symbol $f(x) \sim g(x)$ as $x \to x_0$ denotes asymptotic equality: $\lim_{x\to x_0} \frac{f(x)}{g(x)} = 1$.

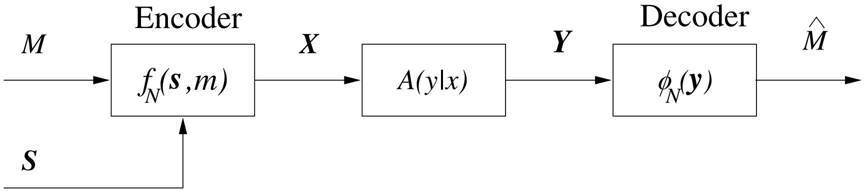


Fig. 1. The blind watermark communication problem

2.1 Mathematical Model

Let $d(x,y) = (x-y)^2$ be the squared-error distortion measure. Referring to Fig. 1, a message m is drawn uniformly from a message set $\mathcal{M}$ and embedded in a length-N host sequence $\mathbf{s} = (s_1, \cdots, s_N)$ using an encoding function $\mathbf{x} = f_N(\mathbf{s}, m)$. The following average-distortion constraint is imposed on f_N:

$$\mathbb{E}\|\mathbf{X} - \mathbf{S}\|^2 = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \int_{\mathbb{R}^N} \|f_N(\mathbf{s}, m) - \mathbf{s}\|^2 p(\mathbf{s}) d\mathbf{s} \leq N D_1.$$

A *memoryless attack channel, subject to distortion* D_2, is a conditional probability density function (pdf) $A(y|x)$, $x, y \in \mathbb{R}$, subject to distortion constraints. The length-N extension of this channel is defined as $A^N(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^N A(y_i|x_i)$. The attack channel is subject to an average-distortion constraint:

$$\mathbb{E}\|\mathbf{Y} - \mathbf{S}\|^2 = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \|\mathbf{y} - \mathbf{s}\|^2 A^N(\mathbf{y}|f_N(\mathbf{s}, m))\, p(\mathbf{s})\, d\mathbf{s}\, d\mathbf{y} \leq N D_2, \tag{1}$$

i.e., distortion is measured with respect to the host.

We require $D_2 \geq D_1$, so that the feasible set of attack channels includes $\mathbf{Y} = \mathbf{X}$ (no attack). The distortions for the information hider and the attacker are equal in this special case.

2.2 Watermarking Capacity

Watermarking capacity is defined as the supremum of all achievable transmission rates, where the supremum is taken over all encoding functions subject to distortion D_1. This capacity is the value of a mutual-information game between the information hider and the attacker [11]. First, the information hider designs a *covert channel* $Q(x, u|s)$, where U is an auxiliary real-valued random variable. The covert channel satisfies the distortion constraint

$$\int\int\int (x-s)^2 Q(x,u|s) p(s)\, dx ds du \leq D_1. \tag{2}$$

Next, the attacker designs an attack channel $A(y|x)$ that satisfies the distortion constraint

$$\int\int\int\int (y-s)^2 A(y|x) Q(x,u|s) p(s)\, ds dx du dy \leq D_2. \tag{3}$$

Let $\mathcal{A}(Q, D_2)$ and $\mathcal{Q}(D_1)$ be the set of channels that satisfy the constraints (3), and (2), respectively. The dependency of $\mathcal{A}(Q, D_2)$ on Q is via the marginal $p(x|s)$. The capacity is given by [11][1]

$$C = \sup_{Q(x,u|s) \in \mathcal{Q}(D_1)} \min_{A(y|x) \in \mathcal{A}(Q,D_2)} J(Q, A) \tag{4}$$

where

$$J(Q, A) = I(U;Y) - I(U;S). \tag{5}$$

2.3 Gaussian Channels

When the host S is Gaussian, the optimal covert channel admits an elegant closed-form solution: X is the output of a Gaussian test channel with distortion D_1, whose input is S. The optimal attack is the Gaussian test channel with distortion level $D_2 - D_1$. The solution is stated in Theorem 1 below, and the capacity-achieving marginal pdf of (S, X, Y) is depicted in Fig. 2. All capacity expressions in this paper are given in terms of the function

$$C_{AWGN}(SNR) = \frac{1}{2} \log(1 + SNR) \tag{6}$$

which is Shannon's capacity formula for the AWGN channel with signal-to-noise ratio equal to SNR. Moreover, all capacity expressions depend on σ_s^2, D_1 and D_2 only via the *watermark-to-noise ratio*

$$\mathrm{WNR} \triangleq \frac{D_1}{D_2} \leq 1 \tag{7}$$

(where the inequality follows from our discussion below (1)), and the *noise-to-host ratio*

$$\mathrm{NHR} \triangleq \frac{D_2}{\sigma_s^2}. \tag{8}$$

Assuming that $D_2 > 0$, the case NHR = 0 corresponds to the limiting case of a Gaussian host pdf with unbounded variance.

Theorem 1. *[12] Assume blind watermarking of a Gaussian host $S \sim \mathcal{N}(0, \sigma_s^2)$.*
(i) *If* $\mathrm{NHR} \geq 1$*, the optimal attack channel is given by $Y = 0$, and capacity is $C = 0$.*
(ii) *If* $\mathrm{NHR} < 1$*, capacity is given by*

$$C(\mathrm{WNR}, \mathrm{NHR}) = C_{AWGN}\left(\frac{\mathrm{WNR}(1 - \mathrm{NHR})}{1 - \mathrm{WNR}}\right). \tag{9}$$

The optimal attack channel $A(y|x)$ is the Gaussian test channel:

$$Y = \frac{1}{\beta}(X + W), \tag{10}$$

[1] This theorem was stated in [12] under the assumption that the decoder knows the attack channel A, however this restriction is now known to be unnecessary.

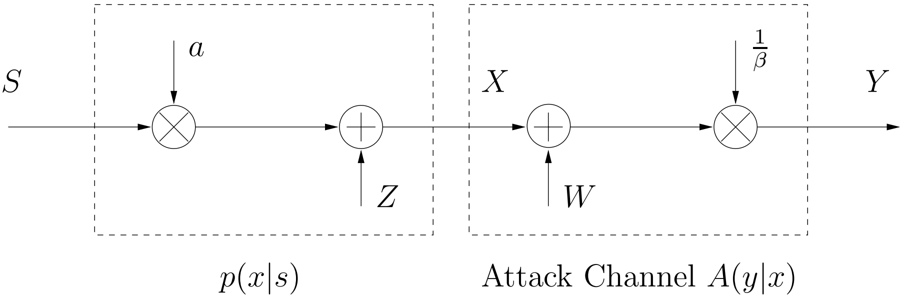


Fig. 2. Minmax-optimal $p(x|s)$ and $A(y|x)$ for i.i.d. Gaussian host data $S \sim \mathcal{N}(0, \sigma_s^2)$ under type-S distortion constraints. Both $p(x|s)$ and $A(y|x)$ are Gaussian test channels.

where $W \sim \mathcal{N}(0, \sigma_w^2)$ *is independent of* X*,*

$$\sigma_w^2 = D_2(1 - \mathrm{WNR})\frac{1 - \mathrm{WNR} * \mathrm{NHR}}{1 - \mathrm{NHR}}, \tag{11}$$

and

$$\beta = \frac{\sigma_x^2}{\sigma_x^2 - (D_2 - D_1)} = \frac{1 - \mathrm{WNR} * \mathrm{NHR}}{1 - \mathrm{NHR}}. \tag{12}$$

The optimal covert channel $Q(x, u|s)$ *is given by*

$$X = aS + Z \tag{13}$$

$$U = \alpha S + Z \tag{14}$$

where $Z \sim \mathcal{N}(0, \sigma_z^2)$ *is independent of* S*,*

$$\sigma_z^2 = aD_1, \quad a = 1 - \mathrm{WNR} * \mathrm{NHR}, \quad \alpha = \frac{\sigma_z^2}{\sigma_z^2 + \sigma_w^2}. \tag{15}$$

Remark 1. For small distortions (NHR → 0), we have $a, \beta \to 1$ and $C \sim C_{AWGN}(\frac{\mathrm{WNR}}{1-\mathrm{WNR}})$. The AWGN attack is asymptotically optimal as NHR → 0.

Remark 2. For fixed WNR, the capacity expression (9) is zero for NHR ≥ 1 and strictly decreasing in NHR for $0 \leq \mathrm{NHR} \leq 1$. Informally speaking, capacity is zero if the host is too weak; and capacity increases with the randomness of S (NHR → 0). Based on the discussion above, the range of nontrivial values for (WNR, NHR) is given by

$$0 \leq \mathrm{NHR} \leq 1, \quad 0 \leq \mathrm{WNR} \leq 1. \tag{16}$$

Remark 3. We may write (9) as $C(\mathrm{WNR}, \mathrm{NHR}) = C_{AWGN}(\mathrm{WNR}_{\mathrm{eff}})$ where

$$\mathrm{WNR}_{\mathrm{eff}} = \frac{\sigma_z^2}{\sigma_w^2} = \frac{\mathrm{WNR}(1 - \mathrm{NHR})}{1 - \mathrm{WNR}}. \tag{17}$$

Therefore $C(\mathrm{WNR}, \mathrm{NHR})$ is the capacity of an AWGN channel with input power σ_z^2 and noise power σ_w^2. The formula can be interpreted by referring to Fig. 2 and recalling Costa's result [21]. The known interference aS does not reduce capacity, and neither, of course, does the known constant β.

Remark 4. In (17), $\mathrm{WNR}_{\mathrm{eff}}$ is a convex, increasing function of WNR. Observe that $\mathrm{WNR}_{\mathrm{eff}} = 0$ for $\mathrm{WNR} = 0$; $\mathrm{WNR}_{\mathrm{eff}} = \mathrm{WNR}$ for $\mathrm{WNR} = \mathrm{NHR}$; and $\mathrm{WNR}_{\mathrm{eff}} = \infty$ for $\mathrm{WNR} = 1$.

Remark 5. The optimal attack when $a = 1$ is the minimum-mean-squared-error (MMSE) estimator of S given X cascaded with a Gaussian test channel. The MMSE operation helps the attacker in reducing the distortion with respect to S, making it possible for the noise source W to have larger variance.

Remark 6. For the optimal choice $a = 1 - \mathrm{WNR} * \mathrm{NHR}$, the MMSE estimator of S given X is X itself. Obviously this choice of a makes the MMSE operation least useful for the attacker.

3 Spread-Spectrum Modulation

Additive SSM is a linear modulation technique, commonly formulated as

$$\mathbf{X} = \mathbf{S} + \mathbf{Z}_m \tag{18}$$

where the vector $\mathbf{Z}_m$ is indexed by the message m to be sent. For weak hosts, a simple but effective enhancement is to attenuate the host prior to embedding [16,17]:

$$\mathbf{X} = a\mathbf{S} + \mathbf{Z}_m \tag{19}$$

where $0 \le a \le 1$ is the attenuation factor. This technique was later called *distortion-compensated SSM* [7]. However, since the attenuation mechanism is fundamentally different from the distortion-compensation mechanism used in QIM, we shall simply refer to (19) as *aSSM.*

Remarkably, optimization of the attenuation factor a against the worst attack in class $\mathcal{A}(Q, D_2)$ results in the same solution as in Fig. 2, hence in the same optimal values of a, σ_z^2, σ_w^2, and β. This follows from [17], where, like here, the cost function is effective SNR at the receiver.

For aSSM, the effective noise power is $a^2\sigma_s^2 + \sigma_w^2$, and the effective signal power is σ_z^2. This results in an effective signal-to-noise ratio

$$\begin{aligned} \mathrm{WNR}_{aSSM} &= \frac{\sigma_z^2}{a^2\sigma_s^2 + \sigma_w^2} \\ &= \frac{\mathrm{WNR} * \mathrm{NHR}\,(1 - \mathrm{NHR})}{1 + (\mathrm{NHR}^2 - 2 * \mathrm{NHR})\mathrm{WNR}} \\ &\le \mathrm{WNR}_{\mathrm{eff}} \end{aligned} \tag{20}$$

(where equality holds only in the trivial cases $NHR = 1$ and $WNR = 0$), and the capacity function

$$C_{AWGN}(\mathrm{WNR}_{aSSM}) < C_{AWGN}(\mathrm{WNR}_{\mathrm{eff}}) = C(\mathrm{WNR}, \mathrm{NHR}). \tag{21}$$

As expected from such a simple linear modulation scheme, aSSM is not capacity-achieving. Conventional SSM is even worse. However, note from (20) and (17) that $\mathrm{WNR}_{aSSM} \sim \mathrm{WNR}_{\mathrm{eff}}$ as NHR → 1. Therefore we may conclude from (21) that aSSM is asymptotically capacity-achieving as NHR → 1 for all values of WNR $\in [0, 1)$. Another way to look at this property follows from (11) and (15): we have

$$\frac{\sigma_w^2}{a^2\sigma_s^2} \sim \frac{1}{1 - \mathrm{NHR}} \to \infty \quad \text{as NHR} \to 1.$$

That is, the attacker's noise W dominates the host signal aS, and the communication model becomes equivalent to the standard AWGN model without side information at the encoder.

4 Improved Spread Spectrum

One may ask whether further improvements on aSSM are possible using Malvar and Florêncio's Improved Spread Spectrum (ISS) method [18], in which different attenuation factors and watermark powers are allocated to different host signal components (different subliminal channels). The fundamental potential advantage of ISS over aSSM resides in the ability to keep the subliminal channels unknown to the attacker; otherwise an analysis similar to that in [12] shows that there is nothing to be gained by such strategy.

A mathematically tractable version of ISS would be the following. Host signal samples $S_1, \cdots, S_N$ are divided into K (secret) groups with size $N_k = \lfloor r_k N \rfloor$, where $1 \leq k \leq K$ and $\sum_{k=1}^K r_k = 1$. For each group a different attenuation factor a_k and watermark power $\sigma_{z,k}^2$ is used, resulting in a per-sample embedding distortion of $D_{1k} = (a_k - 1)^2\sigma_s^2 + \sigma_{z,k}^2$. Define the random variables $\overline{Z}$ and $\overline{X}$ taking values Z_k and X_k respectively, with probability r_k for $1 \leq k \leq K$. The time-average embedding distortion is $D_1 = \sum_{k=1}^K r_k D_{1,k}$, and the variance of $\overline{X}$ (also equal to the time-averaged variance of X) is $\sigma_x^2 = \sum_k r_k \sigma_{x,k}^2$. Similarly, $\overline{Z}$ has variance $\sigma_z^2 = \sum_k r_k \sigma_{z,k}^2$; moreover, $\overline{Z}$ and S are independent. We assume that the attacker knows the joint statistics of $(S, \overline{X})$ but not the subliminal channels and implements a *memoryless* Gaussian channel $Y = (\overline{X} + W)/\beta$ subject to the distortion constraint D_2; W is independent of $\overline{X}$. We may not assume that the second-order statistics of $(S, \overline{X})$ and $(\overline{X}, Y)$ are those of Gaussian test channels.

The capacity function for ISS may be written as

$$C_{ISS}(\mathrm{WNR}, \mathrm{NHR}) = \max_{\sigma_x^2} \min_{\sigma_w^2} \tilde{C}_{ISS}(D_1, \sigma_x^2, \sigma_w^2) \tag{22}$$

where

$$\tilde{C}_{ISS}(D_1, \sigma_x^2, \sigma_w^2) = \max_{\mathbf{r}, \{a_k, \sigma_{z,k}^2\}} \sum_{k=1}^{K} r_k C_{AWGN}\left(\frac{\sigma_{z,k}^2}{{a_k}^2 \sigma_s^2 + \sigma_w^2}\right). \tag{23}$$

The maximization is subject to the constraints

$$\sum_{k=1}^{K} r_k[(a_k-1)^2\sigma_s^2 + \sigma_{z,k}^2] = D_1, \quad \sum_{k=1}^{K} r_k[a_k^2\sigma_s^2 + \sigma_{z,k}^2] = \sigma_x^2.$$

The maximization over $\mathbf{r}$ takes place over the probability simplex. Therefore $\tilde{C}_{ISS}(D_1, \sigma_x^2, \sigma_w^2)$ is the upper convex envelope (with respect to D_1) of the function

$$\tilde{C}(D_1, \sigma_x^2, \sigma_w^2) = \max_{a, \sigma_z^2} C_{AWGN}\left(\frac{\sigma_z^2}{a^2\sigma_s^2 + \sigma_w^2}\right), \quad 0 \le D_1 \le \sigma_s^2, \tag{24}$$

where the maximization is subject to the constraints

$$(a-1)^2\sigma_s^2 + \sigma_z^2 = D_1, \quad a^2\sigma_s^2 + \sigma_z^2 = \sigma_x^2.$$

Therefore the feasible set for (a, σ_z^2) is a singleton. After some simple algebra, we can establish that the function $\tilde{C}(D_1, \sigma_x^2, \sigma_w^2)$ is convex in D_1 for all $D_1 \ge D_1^* = \sigma_x^2 + \sigma_s^2 - 2\sigma_s\sigma_w$ but *concave otherwise*. Hence its upper convex envelope is

$$\tilde{C}_{ISS}(D_1, \sigma_x^2, \sigma_w^2) = \begin{cases} D_1 \frac{\tilde{C}(D_1^*, \sigma_x^2, \sigma_w^2)}{D_1^*} & : \ D_1 < D_1^* \\ \tilde{C}(D_1, \sigma_x^2, \sigma_w^2) & : \ \text{else.} \end{cases} \tag{25}$$

At most two subliminal channels are needed to achieve ISS capacity. Observe the following special cases:

- $\frac{\sigma_w}{\sigma_s} \ge \frac{\sigma_x^2 + \sigma_w^2}{2\sigma_s^2}$: in this case, $D_1^* \le 0$, and $\tilde{C}_{ISS}(D_1, \sigma_x^2, \sigma_w^2) = \tilde{C}(D_1, \sigma_x^2, \sigma_w^2)$. This corresponds to the case of high NHR, with WNR not too close to 1.
- $\frac{\sigma_w}{\sigma_s} \le \frac{\sigma_x^2}{2\sigma_s^2}$: in this case, $D_1^* \ge \sigma_s^2$, and $\tilde{C}_{ISS}(D_1, \sigma_x^2, \sigma_w^2) > \tilde{C}(D_1, \sigma_x^2, \sigma_w^2)$ is in the straight-line regime. This corresponds to the case of low NHR.

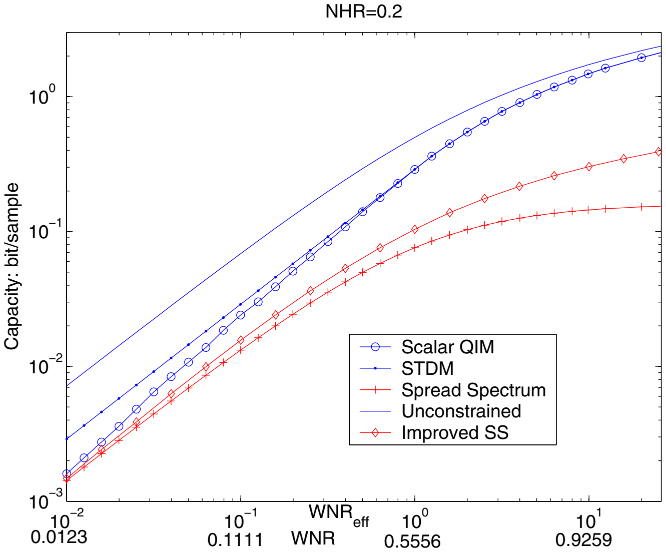
On the other hand, the capacity function for aSSM is given by

$$C_{AWGN}(\text{WNR}_{aSSM}) = \max_{\sigma_x^2} \min_{\sigma_w^2} \tilde{C}(D_1, \sigma_x^2, \sigma_w^2). \tag{26}$$

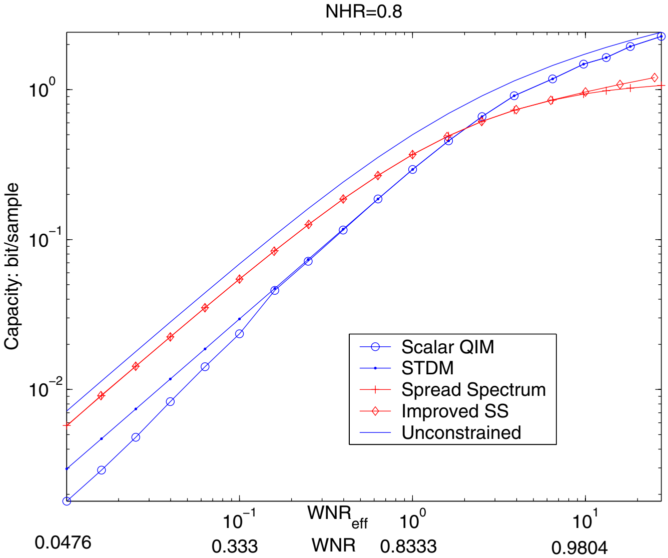
Due to (22) and (25), the aSSM capacity cannnot exceed $C_{ISS}(\text{WNR}, \text{NHR})$. Equality is achieved at high NHR, provided WNR is not too close to 1. A potential advantage of ISS over aSSM[2] appears at low NHR, as illustrated by the numerical results in Fig. 3. Plots are given for a low value of NHR and for a large value of NHR.[3]

[2] Also note that ISS presents dramatic advantages over aSSM in terms of error probability for zero-rate watermarking [20].

[3] Some of the values of (NHR, WNR) used in Fig. 3 are likely to be unrealistic in a practical application. We use them to illustrate the limiting performance of the various schemes considered.



(a) NHR = 0.2



(b) NHR = 0.8.

Fig. 3. Capacity curves: C is plotted on a log scale as a function of $\text{WNR}_{\text{eff}} \geq 0$ (corresponding values of $\text{WNR} \in [0, 1]$ are indicated underneath)

5 Dithered Lattice QIM

Let

Λ = lattice in Euclidean space $\mathbb{R}^L$;
Q = quantization function mapping each point $\mathbf{x} \in \mathbb{R}^L$ to the nearest lattice point in Λ;
$\mathcal{V} = \{\mathbf{x} \in \mathbb{R}^L \; : \; Q(\mathbf{x}) = 0\}$ = Voronoi cell of Λ.

Let M be an integer, and $R = \frac{1}{L}\log_2 M$. Consider the problem of embedding a message $m \in \{0, 1, \cdots, M-1\}$. A rate-$R$ lattice QIM embedding scheme is defined by a set of vectors $\{\mathbf{z}_m, 0 \le m < M\}$, a lattice inflation parameter $0 \le \alpha \le 1$ (*aka* Costa parameter), and the embedding function

$$\mathbf{x} = Q(\alpha\mathbf{s} + \mathbf{z}_m - \mathbf{d}) + (1-\alpha)\mathbf{s} - \mathbf{z}_m + \mathbf{d}. \tag{27}$$

The vector $\mathbf{d}$ in (27) is an external dither vector that is randomized uniformly over $\mathcal{V}$ and independent of $\mathbf{s}$ and m, and is known to the decoder. Such randomization achieves two purposes: (1) it facilitates the proof of capacity theorems [4] and error exponent analyses [**?**,**?**], and (2) it provides a certain level of security against attackers that are not limited to additive-noise attacks. In the remainder of this section, we assume that $\mathbf{d}$ satisfies the statistical model above. This makes the *self-noise* due to quantization uniformly distributed over the scaled Voronoi cell $(1-\alpha)\mathcal{V}$ and independent of $\mathbf{s}$ and m.

A natural idea in our problem with nonadditive attacks is to apply lattice QIM to the attenuated signal aS, resulting in the embedding formula

$$\mathbf{x} = Q(\alpha a\mathbf{s} + \mathbf{z}_m - \mathbf{d}) + (1-\alpha)a\mathbf{s} - \mathbf{z}_m + \mathbf{d}. \tag{28}$$

Analogously to aSSM in (19), this scheme could be termed *aQIM.* The maximum achievable rate for L-dimensional aQIM is given by

$$\tilde{C}_L(\text{WNR}) = C_L(\text{WNR}_{\text{eff}})$$

where

$$C_L(\text{WNR}_{\text{eff}}) \triangleq \max_{0\le\alpha\le1} \max_{\Lambda} \max_{p_{\mathbf{Z}}} I(\mathbf{Z}; \tilde{\mathbf{Y}}) \tag{29}$$

is the capacity function for the Erez-Zamir scheme. In (29), $p_{\mathbf{Z}}$ is a pdf over the Voronoi cell $\mathcal{V}$ of Λ, and

$$\tilde{\mathbf{Y}} = \alpha\beta\mathbf{Y} \bmod \Lambda = \alpha\beta\mathbf{Y} - Q(\alpha\beta\mathbf{Y}) \tag{30}$$

is the output of the lattice-reduction step at the decoder. The capacity formula (29) *can be* obtained by analyzing the MAN *vector channel* of Fig. 4.[4] The noise $\mathbf{V}$ in this channel is the sum (mod Λ) of the self-noise and the scaled attacker's noise, $\alpha\mathbf{W} \sim \mathcal{N}(0, \alpha^2\sigma_w^2 I_L)$.

From Remark 3 in Sec. 2, we immediately obtain the following result.

[4] The method of proof used by Erez and Zamir [4] is somewhat different.

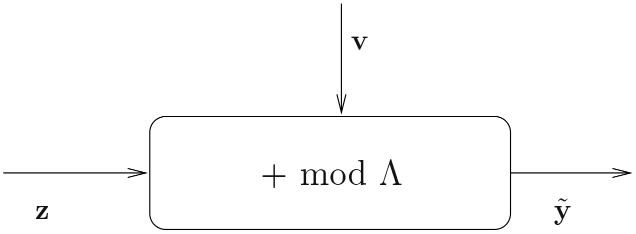


Fig. 4. Modulo Additive Noise (MAN) channel for lattice QIM

Proposition 1. *The aQIM scheme (28) achieves the unconstrained capacity bound (9), in the limit as the lattice dimension L tends to infinity.*

The capacity-achieving p_Z in (29) is uniform over $\mathcal{V}$ [4]. The sequence $C_L(\mathrm{WNR}_{\mathrm{eff}})$ is nondecreasing in L and converges to the unconstrained capacity limit $C_{AWGN}(\mathrm{WNR}_{\mathrm{eff}})$ as $L \to \infty$. It is remarkable that the lattice-reduction step (30), which is information-lossy, does not cause a loss of capacity. Therefore any substantial improvement to QIM would have to be restricted to low-dimensional QIM.

Consider the two extreme values of NHR in (16), and fix WNR.

- For NHR = 0 we have $\mathrm{WNR}_{aSSM} = 0$; hence scalar QIM outperforms SSM.
- As NHR → 1, we have $\mathrm{WNR}_{aSSM} \sim \mathrm{WNR}_{\mathrm{eff}}$; both tend to zero.

Fig. 3 compares capacity functions for scalar QIM in (28) and aSSM in (19) to the unconstrained capacity formula (9). These plots illustrate the superiority of scalar QIM over aSSM for low NHR, but also the high performance of aSSM for large NHR. We reemphasize that the advantage of aSSM over scalar QIM in this case is due to the low dimensionality of the lattice QIM scheme used.

6 Time-Shared Lattice QIM

The scalar QIM capacity function $C_1(\mathrm{WNR}_{\mathrm{eff}})$ is nonconvex. Therefore time-sharing can be used to improve capacity performance [5,19]. In a time-shared scheme, transmission takes place during a fraction τ of the time; the transmission power during that time is boosted by a factor of $\frac{1}{\tau}$. The effect of time-sharing is to convexify the capacity function $C_1(\mathrm{WNR}_{\mathrm{eff}})$. The resulting (improved) capacity function is given by[5]

$$\begin{aligned}\tilde{C}_{STDM}(\mathrm{WNR}) &= C_{STDM}(\mathrm{WNR}_{\mathrm{eff}}) \\ &= \begin{cases} \mathrm{WNR}_{\mathrm{eff}} \, \frac{C_1(\mathrm{WNR}^*)}{\mathrm{WNR}^*} & : \; 0 \le \mathrm{WNR}_{\mathrm{eff}} < \mathrm{WNR}^* \\ C_1(\mathrm{WNR}_{\mathrm{eff}}) & : \; \mathrm{WNR}_{\mathrm{eff}} \ge \mathrm{WNR}^* \end{cases}\end{aligned} \tag{31}$$

[5] Recall our assumption that the attacker does not know the subliminal channels and sticks to the memoryless attack of Theorem 1.

i.e., improvements are obtained at all $\mathrm{WNR}_{\mathrm{eff}}$ below a critical value $\mathrm{WNR}^* \approx 1$; the corresponding critical value of WNR is obtained from (17) as $\frac{\mathrm{WNR}^*}{\mathrm{WNR}^*+1-\mathrm{NHR}}$. We also have $C_1(\mathrm{WNR}^*) \approx 0.3$ bits ≈ 0.2 nats. The straight-line portion of the curve (31) is obtained by varying τ from 0 to 1. Time-shared scalar QIM was introduced by Chen and Wornell under the name Spread Transform Dither Modulation (STDM) [1] and further studied by Eggers *et al* [2].

Due to (17), the capacity expression (31), viewed as a function of WNR and measured in nats, has slope at the origin equal to $(1-\mathrm{NHR})\,\frac{C_1(\mathrm{WNR}^*)}{\mathrm{WNR}^*}$. The slope of the spread-spectrum capacity function (21) at $\mathrm{WNR}=0$ is equal to

$$\left.\frac{d(\mathrm{WNR}_{aSSM})}{d(\mathrm{WNR})}\right|_{\mathrm{WNR}=0} = \frac{1}{2}\mathrm{NHR}\,(1-\mathrm{NHR}).$$

Therefore a necessary condition for time-shared scalar QIM to outperform SSM *at all* WNR's is

$$\frac{C_1(\mathrm{WNR}^*)}{\mathrm{WNR}^*} \geq \frac{1}{2}\mathrm{NHR}. \qquad (32)$$

i.e., $\mathrm{NHR} \leq 0.4$. If the condition above is violated, then aSSM outperforms time-shared scalar QIM at low WNR's, as illustrated in Fig. 3.

This derivation carries to the higher-dimensional case. Lattice QIM can be improved at low WNR's using time-sharing, but WNR^* tends to 0 as $L \to \infty$[6]; therefore $\lim_{L\to\infty} \frac{C_L(\mathrm{WNR}^*)}{\mathrm{WNR}^*} = \frac{1}{2}$. Moreover, equation (32) holds with C_L in place of C_1, and thus in the limit as $L \to \infty$, lattice QIM outperforms SSM *at all* NHR and WNR, in agreement with our earlier analysis.

7 Nondithered Lattice QIM

For scalar QIM subject to AWGN attacks, numerical experiments by Pérez-Freire and Pérez-González [7,8] have revealed possible improvements in communication performance if no external dither is used in (27).[7]

Here we comment upon this interesting result from an analytical perspective. The mutual information $I(\mathbf{Z};\tilde{\mathbf{Y}})$ for lattice QIM can be written as

$$\begin{aligned} I(\mathbf{Z};\tilde{\mathbf{Y}}) &= \int_{\mathcal{V}} p_{\mathbf{D}}(\mathbf{d}) I(\mathbf{Z};\tilde{\mathbf{Y}}|\mathbf{D}=\mathbf{d})d\mathbf{d} \\ &\leq \max_{\mathbf{d}\in\mathcal{V}} I(\mathbf{Z};\tilde{\mathbf{Y}}|\mathbf{D}=\mathbf{d}) \\ &= \max_{\mathbf{d}\in\mathcal{V}} I(\mathbf{Z}-\mathbf{d};\tilde{\mathbf{Y}}|\mathbf{D}=0) \qquad (33) \end{aligned}$$

[6] By convexity of the limiting capacity function $C_{AWGN}(\mathrm{WNR}_{\mathrm{eff}})$.

[7] Additionally, Pérez-Freire and Pérez-González [7,8] discovered that additional improvements – albeit minor ones – are obtained if the lattice-reduction step (30) is omitted at the decoder. The existence of an improvement follows from the data-processing inequality [15]: $I(\mathbf{Z};\tilde{\mathbf{Y}}) \leq I(\mathbf{Z};\mathbf{Y})$. As discussed in the Introduction, the improvement vanishes for higher-order QIM, as $L \to \infty$.

where the last line follows from the definition (27) of the dithered QIM scheme. In the Erez-Zamir scheme, $p_{\mathbf{D}}$ is chosen to be uniform over $\mathcal{V}$; as we know, this choice is asymptotically optimal as $L \to \infty$. It is however clear that if one uses $p_{\mathbf{D}}$ as an additional variable to be optimized, one will do at least as well as the Erez-Zamir scheme. Specifically, the upper bound in (33) is achieved by a mass distribution $p_{\mathbf{D}}$ located at some $\mathbf{d}^* \in \mathcal{V}$. Optimizing the left side of (33) over $p_{\mathbf{Z}}$, we obtain

$$\begin{aligned} \max_{p_{\mathbf{Z}}, p_{\mathbf{D}}} I(\mathbf{Z}; \tilde{\mathbf{Y}}) &= \max_{p_{\mathbf{Z}}} \max_{\mathbf{d} \in \mathcal{V}} I(\mathbf{Z} - \mathbf{d}; \tilde{\mathbf{Y}} | \mathbf{D} = 0) \\ &= \max_{p_{\mathbf{Z}}} I(\mathbf{Z}; \tilde{\mathbf{Y}} | \mathbf{D} = 0), \end{aligned}$$

i.e., the cost function is maximized when no external dither is used! Note that the maximizing $p_{\mathbf{Z}}$ above is not necessarily uniform over $\mathcal{V}$ (as was the case in the Erez-Zamir scheme). For any L, α, and Λ, the nondithered design improves over the Erez-Zamir design with uniform dither; however, as mentioned above, the performance gap vanishes for large L.

For small L, the performance gap may be substantial. Indeed Erez and Zamir showed that at high $\mathrm{WNR}_{\mathrm{eff}}$, the gap to capacity for their scheme is equal to the shaping gain of lattice VQ (about 1.53 dB in the scalar case, $L = 1$). In the case NHR → 1, aSSM is capacity-achieving and therefore outperforms (by about 1.53 dB when $L = 1$) dithered L-dimensional QIM.[8] The following proposition shows that nondithered aQIM does *much better*.

Proposition 2. *The capacity function of aSSM cannot exceed that of nondithered aQIM for any value of* NHR*,* WNR*, and* L*.*

Sketch of the proof: To prove the claim, it suffices to consider the scalar QIM case ($L = 1$) and identify a particular value of the lattice inflation parameter α and of the quantizer step size Δ, as well as a pdf p_Z, such that nondithered QIM and aSSM have the same capacity performance. Let $\Delta \to \infty$ and choose $\alpha = 0$ and $p_Z = \mathcal{N}(0, \sigma_z^2)$. Then $Q(a\alpha S + Z) = 0$ with probability tending to 1, and (28) becomes $X = aS - Z$ (with probability one), i.e., coincides almost surely with aSSM. The actual proof is based on the continuity of mutual information with respect to variational norm. □

From the proposition above, one could numerically optimize α, Δ and p_Z to devise a scalar QIM scheme that outperforms both the aSSM and dithered aQIM schemes. We have not attempted such costly optimization, simply noting that time-sharing between aSSM and aQIM (with time-sharing parameter determined by the values of WNR and NHR) achieves the convex hull of the aSSM and aQIM capacity curves, which may be good enough for practical purposes. Similarly to the proposition above, we also have

Proposition 3. *The capacity function of ISS cannot exceed that of nondithered STDM for any value of* NHR *and* WNR*.*

[8] Some care is needed about the order in which asymptotics are taken. To have both NHR → 1 and $\mathrm{WNR}_{\mathrm{eff}} \to \infty$, we need that $1 - \mathrm{WNR} \ll 1 - \mathrm{NHR}$.

8 Discussion

We have considered Gaussian host signals and studied the effects of NHR $\in [0,1]$ on the capacity function of constrained watermark embedding schemes, allowing nonadditive attacks with bounded squared distortion. For unconstrained schemes the worst attack is known to be the Gaussian test channel. When NHR $\to 0$ (host signal whose variance tends to infinity), additive attacks are optimal. For NHR $= 1$, the compression attack $Y = 0$ is feasible, and capacity is zero. We have introduced the aQIM scheme, which is a simple variation on the Erez-Zamir scheme [4], and compared its performance with that of the aSSM linear modulation scheme [16,17]. Our results are summarized as follows.

1. Prop. 1: At all NHR's, L-dimensional aQIM with uniform dither and lattice reduction at the decoder is asymptotically capacity-achieving, as $L \to \infty$.
2. In the extreme case NHR $\to 1$, host-signal interference is weak, and attacker's noise dominates at the decoder. aSSM is asymptotically capacity-achieving and outperforms *low-dimensional* aQIM schemes with uniform dither.
3. For any finite choice of L, the aQIM scheme can be improved by eliminating the external dither (and keeping the lattice reduction step at the decoder); the improvement vanishes as $L \to \infty$.
4. Prop. 2: The nondithered aQIM scheme outperforms aSSM [16,17] for all values of L, WNR, and NHR.
5. Prop. 3: The nondithered STDM scheme outperforms ISS [18] for all values of L, WNR, and NHR.

Clearly, the potential for improving the attacker's performance exists in the form of non-Gaussian strategies and strategies with memory.

Acknowledgements. The authors thank Tie Liu and the reviewers for helpful comments and suggestions.

References

1. B. Chen and G. W. Wornell, "Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding," *IEEE Trans. on Information Theory*, Vol. 47, No. 4, pp. 1423—1443, May 2001.
2. J. J. Eggers, R. Bäuml, R. Tzschoppe and B. Girod, "Scalar Costa Scheme for Information Embedding," *IEEE Transactions on Signal Processing*, Vol. 51, No. 4, pp. 1003—1019, Apr. 2003.
3. R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested Linear/Lattice Codes for Structured Multiterminal Binning," *IEEE Trans. Information Theory*, Vol. 48, No. 6, pp. 1250—1276, June 2002.
4. U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+SNR)$ on the AWGN Channel with Lattice Encoding and Decoding," *IEEE Trans. on Information Theory*, Vol. 50, No. 10, pp. 2293—2314, Oct. 2004.
5. U. Erez and S. ten Brink, "Approaching the dirty paper limit for canceling known interference," *Proc. Allerton Conf.*, Monticello, IL, Sep. 2003.

6. O. Koval, S. Voloshynovskiy, and F. Pérez-González, “Quantization-Based Watermarking Performance Improvement Using Host Statistics: AWGN Attack Case,” *Proc. ACM Multimedia and Security Workshop*, Magdeburg, Germany, Sep. 2004.
7. L. Pérez-Freire, F. Pérez-González, and S. Voloshynovskiy, “Revealing the True Achievable Rates of Scalar Costa Scheme,” *Proc. IEEE Multimedia Signal Proc. Workshop*, Siena, Italy, Sep.-Oct. 2004.
8. L. Pérez-Freire and F. Pérez-González, “Spread-Spectrum vs Quantization-Based Data Hiding: Misconceptions and Implications,” *Proc. SPIE*, San Jose, CA, Jan. 2005.
9. I. J. Cox, J. Killian, F. T. Leighton and T. Shamoon, “Secure Spread Spectrum Watermarking for Multimedia,” *IEEE Trans. Image Proc.*, Vol. 6, No. 12, pp. 1673—1687, Dec. 1997.
10. J. A. O’Sullivan, P. Moulin and M. Ettinger, “Information Theoretic Analysis of Steganography,” *Proc. Int. Symp. on Information Theory (ISIT’98)*, Boston, MA, Aug. 1998.
11. P. Moulin and J. A. O’Sullivan, “Information–Theoretic Analysis of Information Hiding,” *IEEE Trans. on Information Theory*, Vol. 49, No. 3, pp. 563-593, March 2003.
12. P. Moulin and M. K. Mıhçak, “The Parallel-Gaussian Watermarking Game,” *IEEE Trans. on Information Theory*, Vol. 50, No. 2, pp. 272-289, Feb. 2004.
13. P. Moulin and M. K. Mıhçak, “A Framework for Evaluating the Data-Hiding Capacity of Image Sources,” *IEEE Trans. on Image Processing*, Vol. 11, No. 9, pp. 1029–1042, Sep. 2002.
14. M. K. Mıhçak and P. Moulin, “Information-Embedding Codes Matched to Local Gaussian Image Models,” *Proc. IEEE Int. Conf. on Image Processing*, Rochester, NY, Sep. 2002.
15. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, 1991.
16. P. Moulin and A. Ivanović, “Game-Theoretic Analysis of Watermark Detection,” *Proc. IEEE Int. Conf. on Image Processing*, Thessaloniki, Greece, Oct. 2001.
17. P. Moulin and A. Ivanović, “The Zero-Rate Spread-Spectrum Watermarking Game,” *IEEE Trans. on Signal Processing*, Vol. 51, No. 4, pp. 1098-1117, Apr. 2003.
18. H. Malvar and D. Florêncio, “Improved Spread Spectrum: A New Modulation Technique for Robust Watermarking,” *IEEE Trans. Signal Processing*, Vol. 51, no. 4, pp. 898—905, April 2003.
19. F. M. J. Willems, “On Gaussian Channels with Side Information at the Transmitter,” *Proc. 9th Symp. on Information Theory in the Benelux*, Enschede, The Netherlands, pp. 129—135, May 1988.
20. T. Liu and P. Moulin, “Error exponents for watermarking game with squared-error constraints,” *Proc. Int. Symp. Information Theory*, Yokohama, Japan, July 2003.
21. M. Costa, “Writing on Dirty Paper,” *IEEE Trans. on Information Theory*, Vol. 29, No. 3, pp. 439—441, May 1983.