Block QIM Watermarking Games *

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

Anil Kumar Goteti
Qualcomm Inc.
675 Campbell Technology Pkwy
Campbell, CA 95008, USA
agoteti@qualcomm.com

August 31, 2005. Revised, January 30 and May 3, 2006

Abstract

While binning is a fundamental approach to blind data embedding and watermarking, an attacker may devise various strategies to reduce the effectiveness of practical binning schemes. The problem analyzed in this paper is design of worst-case noise distributions against L-dimensional lattice Quantization Index Modulation (QIM) watermarking codes. The cost functions considered are (1) probability of error of the maximum-likelihood decoder, and (2) the more tractable Bhattacharyya upper bound on error probability, which is tight at low embedding rates. Both problems are addressed under the following constraints on the attacker's strategy: the noise is independent of the marked signal, blockwise memoryless with block length L, and may not exceed a specified quadratic-distortion level. The embedder's quadratic distortion is limited as well. Three strategies are considered for the embedder: optimization of the lattice inflation parameter (*aka* Costa parameter), dithering, and randomized lattice rotation. Critical in this analysis are the symmetry properties of QIM nested lattices and convexity properties of probability of error and related functionals of the noise distribution. We derive the minmax optimal embedding and attack strategies and obtain explicit solutions as well as numerical solutions for the worst-case noise. The role of the attacker's memory is investigated; in particular, we demonstrate the remarkable effectiveness of impulsive-noise attacks as L increases. The formulation proposed in this paper is also used to evaluate the capacity of lattice QIM under worst-noise conditions.

Keywords: watermarking, data hiding, quantization index modulation, detection theory, game theory, random codes, error exponents, capacity, convex optimization.

*Work supported by NSF under grants CCR 02-08809 and CCR 03-25924, presented in part at ICIP in Singapore, Oct. 2004, and at ICIP in Genoa, Italy, Sep. 2005.

1 Introduction

A variety of blind watermarking and data hiding schemes have been developed over the last ten years. Much attention has been focused on *quantization index modulation* (QIM) methods, which are information-theoretic binning schemes and are relatively easy to implement [1]–[10]. The most important property of binning schemes is their ability to approach fundamental communication limits such as the maximum rate of reliable transmission (data-hiding capacity) for a given family of attack channels and the error exponents (the rate of decay of error probability) at rates below capacity.

Different families of attack channels have been studied in [1]–[10], corresponding to various degrees of generality in the channel model. This includes a single additive white Gaussian noise (AWGN) channel, the family of all additive-noise channels subject to a squared-error distortion constraint, and the family of all channels subject to a squared-error distortion constraint. Most models assume the use of memoryless channels; in other models, this restriction is relaxed. For the watermarking games of [9, 10], in which the host signal is Gaussian and squared-error distortion constraints are imposed on the embedder and the attacker, the capacity-achieving distributions turn out to be Gaussian.

When *no structure* is imposed on the quantizers, random coding methods have been used to prove the achievability of capacity and error exponents [6]–[10]. Such methods are unpractical, but methods based on lattice vector quantization (henceforth referred to as lattice QIM) are more manageable. For AWGN channels, Erez and Zamir [6] have proven that such structural constraints on the quantizers cause no loss of capacity: capacity is achievable asymptotically as lattice dimension tends to infinity. Surprisingly, a similar result applies to error exponents at high rates [8]. However, there is always a performance loss when low-dimensional lattices are used. For scalar QIM [2, 5], the lattice dimension is equal to one, and the watermark-to-noise ratio (WNR) penalty for achieving a target rate is typically between 1.5 and 3 dB.

This paper is about achievable error probabilities for finite-dimensional lattice QIM methods at low embedding rates. Our study includes the popular scalar and hexagonal QIM methods as special cases. We consider the class of additive-noise channels subject to a squared-error distortion constraint, and various memory constraints. Detection-theoretic functionals of the noise probability density function (pdf) are formulated and used as cost functions in a game between the watermark embedder and the attacker. The strategies to be optimized by the embedder include selection of the QIM lattice inflation parameter (*aka* Costa parameter) as well as randomization strategies. Fundamental to our study are the convexity properties of the cost functions used. Our methods are applied to error-probability, Bhattacharyya, and mutual-information functionals of the noise pdf.

Part of this study was reported in the second author's M.S. thesis [11]. Related work, conducted independently of ours, includes Pérez-González [12] and Vila-Forcén *et al.* [13], where the worst noise pdf against scalar QIM was sought using a nonlinear optimization algorithm; and Tzschoppe *et al.* [14], where the worst noise pdf is sought within a nonparametric family, and obtained using an elegant application of the Blahut-Arimoto algorithm. In both [12] and [14], the cost functional was mutual information. In [13], the cost functional was probability of error based on a single received sample, and a minimum-distance decoder.

The general problem of design and analysis of worst-case additive-noise distributions appears frequently in the communications literature, e.g. see the analysis of [15] for worst-noise in binary-

input channels. The mathematical structure of our optimization problems is however quite different from those in [15] because of the fundamentally different nature of the channel in the QIM problem: the channel introduces modulo-lattice additive noise (MLAN). As we shall see, the resulting worst-case distributions differ significantly from those in [15].

This paper is organized as follows. Sec. 2 states our block coding model, and Sec. 3 specializes it to the case of lattice QIM. Sec. 4 defines the lattice maximum-likelihood (ML) decoder and sets up the game between the watermark embedder and the attacker. Sec. 5 solves the game for a simple but insightful special case: scalar QIM with binary alphabets, subject to memoryless attacks. Analytical solutions are derived, and the asymptotic optimality of impulsive-noise attacks at high watermark-to-noise ratios is established. Sec. 6 treats the more general case of a L-dimensional lattice subject to blockwise-memoryless noise, and designs optimal attacks against the QIM code. To counter such attacks, a randomized rotation strategy is proposed for the watermark embedder in Sec. 7; the worst-case noise distribution is now isotropic. Sec. 8 extends these results to *spread transform dither modulation* (STDM) block codes, using a similar randomized rotation strategy: the host signal is projected onto a secret lower-dimensional subspace prior to embedding. The STDM case is more amenable to analytical exploration, and we present closed-form expressions for decoding performance under isotropic impulsive-noise attacks. In Sec. 9, we study the capacity limits of lattice QIM subject to the worst blockwise memoryless noise. Finally, conclusions are presented in Sec. 10.

2 Mathematical Model

Notation: we denote random variables by uppercase letters and individual realizations by lowercase letters. Boldface letters are used to represent sequences and vectors. The Euclidean norm of a vector $\mathbf{x}$ is denoted by $\|\mathbf{x}\|$. The pdf of a random variable X is denoted by p_X. The volume of a set $\mathcal{V} \subset \mathbb{R}^L$ is denoted by $|\mathcal{V}| \triangleq \int_{\mathcal{V}} d\mathbf{x}$, the indicator function of a set $\mathcal{V}$ by $1_{\{\mathbf{x}\in\mathcal{V}\}}$, the Dirac impulse by $\delta(\cdot)$, and the mathematical expectation operator by $\mathbb{E}(\cdot)$. Finally given two functions $f(x)$ and $g(x)$, we write $f(x) \sim g(x)$ (resp. $f(x) \ll g(x)$) as $x \to x_0$ if the ratio $\frac{f(x)}{g(x)}$ tends to 1 (resp. 0) as $x \to x_0$.

Let q and k be integers. (We use $q = 2$ and $q = 3$ as examples throughout this paper). Referring to Fig. 1, we wish to embed a message $m \in \mathcal{M} \triangleq \{1, 2, \cdots, q^k\}$ in a length-n host signal sequence, using a *two-stage code*. The host sequence is subdivided into $n_B \geq k$ blocks of length $L = \frac{n}{n_B}$ each. We denote by $\mathbf{s}(i) = \{s_1(i), \ldots, s_L(i)\} \in \mathbb{R}^L$ the i^{th} host signal block ($1 \leq i \leq n_B$).

In the first stage, $m \in \mathcal{M}$ is mapped to a q-ary sequence (codeword) $\mathbf{c} \in \{0, 1, \cdots, q-1\}^{n_B}$ using an error-correction code (ECC) with rate k/n_B. The ECC will be referred to as the *outer code*. In the second stage, an embedding function $F : \mathbb{R}^L \times \{0, \cdots, q-1\} \times \Theta \to \mathbb{R}^L$ is applied to each block. The i-th marked block is given by $\mathbf{x}(i) = F(\mathbf{s}(i), \mathbf{c}(i), \theta(i))$, where $\theta(i)$ is a secret key (with alphabet Θ) shared with the decoder. The per-sample, mean-squared distortion of the host signal due to embedding is given by $D_1 = \frac{1}{L}\mathbb{E}\|\mathbf{X} - \mathbf{S}\|^2$.

The rate of the two-stage code is given by

$$R \triangleq \frac{1}{n} \log_2 |\mathcal{M}| = \frac{k}{n} \log_2 q \quad \text{bit/sample.} \tag{2.1}$$

The case $k = n_B$ corresponds to an *uncoded scheme* (high payload) in which no ECC is used in the first stage. The other extreme case, $k = 1$, corresponds to a low-payload scheme in which $|\mathcal{M}| = q$ and the ECC could be, for instance, a simple repetition code.

The choice of the attack models and performance metrics used in this paper is based on the assumption that the code rate is low and that the ECC is suitably randomized. Such would be the case, for instance, of random expurgated codes [16, 17] and random constant-composition codes [18].

The attacker chooses a noise pdf $p_{\mathbf{W}}$ that satisfies the expected-distortion constraint

$$\frac{1}{L}\int_{\mathbb{R}^L} \|\mathbf{w}\|^2 p_{\mathbf{W}}(\mathbf{w}) d\mathbf{w} \leq D_2 \tag{2.2}$$

and uses this pdf to generate independent and identically distributed (i.i.d.) L-dimensional random vectors $\mathbf{W}(i), 1 \leq i \leq n_B$, to the marked blocks. Each $\mathbf{W}(i)$ is independent of $\mathbf{X}(i)$. The quantity $\mathrm{WNR} \triangleq D_1/D_2$ is referred to as the *watermark-to-noise ratio*.

The decoder observes the resulting degraded blocks $\mathbf{y}(i) = \mathbf{x}(i) + \mathbf{w}(i)$, and outputs a decision $\hat{m} = \Phi(\{\mathbf{y}(i), \theta(i), 1 \leq i \leq n_B\}) \in \mathcal{M}$, where Φ denotes the decoding function.

Note that our attack model allows for dependencies between the L components of $\mathbf{W}$, but not across blocks. If the ECC were not randomized, the attacker's performance could be improved by exploiting the structure of the ECC. Due to our assumption that the ECC is randomized, this strategy is not considered in this paper.

3 Lattice QIM

Each symbol $\mathbf{c}(i)$ may be embedded into the host signal block $\mathbf{s}(i)$, $1 \leq i \leq n_B$, using various methods [7]. Of interest in this paper are Chen and Wornell's QIM method, their STDM method [1, 2], and more generally, *nested lattice codes* [3, 4, 6, 7]. We briefly review nested lattice codes and introduce the notation.

A lattice Λ in L-dimensional Euclidean space is defined as a set of points in $\mathbb{R}^L$ such that $\mathbf{x} \in \Lambda$ and $\mathbf{y} \in \Lambda$ implies $\mathbf{x} + \mathbf{y} \in \Lambda$ and $\mathbf{x} - \mathbf{y} \in \Lambda$, which equips Λ with the structure of an additive subgroup of $\mathbb{R}^L$ [20].

A nested lattice code consists of a *coarse lattice* Λ and a *fine lattice* Λ_f. The coarse lattice Λ is a sublattice of Λ_f. The fine lattice Λ_f may be decomposed as the union of q *cosets* of Λ:

$$\Lambda_c = \mathbf{z}_c + \Lambda, \quad c \in \{0, 1, \cdots, q-1\},$$

where $\mathbf{z}_c \in \Lambda_f$. The choice of $\mathbf{z}_c$ is nonunique, but it is convenient to choose a minimum-norm vector, in which case $\mathbf{z}_c$ is termed *coset leader* of Λ_c. The set

$$\mathcal{C} = \Lambda_f/\Lambda = \{\Lambda_c,\, c = 0, 1, \cdots, q-1\} \tag{3.1}$$

carries itself a group structure and is termed the *quotient group* of Λ_f by Λ. $\mathcal{C}$ may be efficiently represented by the coset leaders $\mathbf{z}_0 = 0$ and $\mathbf{z}_1, \cdots, \mathbf{z}_{q-1}$. Next we define

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

Finally, a *lattice inflation parameter* α ("Costa parameter") is introduced to control the amount of distortion compensation introduced in the embedding. The host vector $\mathbf{s} \in \mathbb{R}^L$ is marked using the dithered-QIM formula:

$$\mathbf{x} = F(\mathbf{s}, c, \mathbf{d}) \triangleq \mathsf{Q}(\alpha\mathbf{s} - \mathbf{z}_c - \mathbf{d}) + (1-\alpha)\mathbf{s} + \mathbf{z}_c + \mathbf{d}, \tag{3.2}$$

where $\mathbf{d}$ is an *external dither sequence*, randomized over $\mathcal{V}$, and known to the embedder and the decoder but not to the attacker. The (second stage) watermarking code is the triple $(\alpha, \Lambda, \Lambda_f)$. The motivation for using the randomized dither vector $\mathbf{d}$ in the QIM embedding formula (3.2) is twofold: (i) improve security against surgical attacks, and (ii) obtain a tractable model for the self-noise. The following examples of QIM watermarking codes will be used throughout this paper [1].

Example 1 (Chen-Wornell scheme). Let $\Lambda = \Delta\mathbb{Z}^L$ be the scaled L-dimensional cubic lattice and $\Lambda_f = \Lambda \bigcup \Lambda + (\frac{\Delta}{2}, \cdots, \frac{\Delta}{2})$ be $\frac{\Delta}{2}$ times the so-called checkerboard, or D_L, lattice [20]. Here $\mathbf{z}_0 = (0, \cdots, 0)$, $\mathbf{z}_1 = (\frac{\Delta}{2}, \cdots, \frac{\Delta}{2})$, $\mathcal{C} = \{\mathbf{z}_0, \mathbf{z}_1\}$ is isomorphic to the group with two elements, and $q = |\mathcal{C}| = 2$. The lattice D_2 is also known as the quincunx lattice, see Fig. 2(a).

Example 2. Take $L = 2$ and let Λ be the hexagonal A_2 lattice scaled by Δ, and Λ_f be the union of A_2 rotated by $\frac{\pi}{3}$ and scaled by $\frac{\Delta}{\sqrt{3}}$ and A_2 itself, see Fig. 2(c). Then $\mathcal{C}$ is the cyclic group of order three, and $q = 3$.

Since $\mathbf{d}$ is uniformly distributed over $\mathcal{V}$, the quantization noise $\mathbf{e} \triangleq \mathsf{Q}(\mathbf{u} - \mathbf{d}) - \mathbf{u} + \mathbf{d}$ is also uniformly distributed over $\mathcal{V}$ and is independent of $\mathbf{u}$, for any random vector $\mathbf{u}$ [21]. The mean-squared distortion due to embedding is

$$D_1 = \frac{1}{L} E\|\mathbf{e}\|^2 = \frac{1}{L}\frac{1}{|\mathcal{V}|}\int_{\mathcal{V}} \|\mathbf{e}\|^2 d\mathbf{e}. \tag{3.3}$$

Neither distortion nor decoding performance is affected if the code representers $\mathbf{z}_0, \cdots, \mathbf{z}_{q-1}$ in (3.2) are shifted by an arbitrary amount, i.e., if we allow $\mathbf{z}_0 \neq 0$. For instance, in Example 1 above, one may prefer to use $\mathbf{z}_1 = -\mathbf{z}_0 = (\frac{\Delta}{4}, \cdots, \frac{\Delta}{4})$.

As described in Sec. 2, the attacker is subjected to a power constraint (2.2). For a given lattice Λ, the embedder selects α to optimize the performance of the system against the attacks. The attacker knows the watermarking code and optimizes the noise pdf $p_{\mathbf{W}}$. We call this noise pdf the worst attack.

Modulo Operations. The mod Λ operation on any $\mathbf{x} \in \mathbb{R}^L$ is defined as $\mathbf{x} \mod \Lambda \triangleq \mathbf{x} - \mathsf{Q}(\mathbf{x}) \in \mathcal{V}$. The shorthand $\oint_\Omega \triangleq \int_{\Omega \bmod \Lambda}$ denotes integration over a set $\Omega \subset \mathbb{R}^L$, folded into $\mathcal{V}$ using the modΛ operation. Given a function $p : \mathcal{V} \to \mathbb{R}$, we define its $\mathcal{V}$-periodic extension as $\overset{\circ}{p}(\mathbf{x}) \triangleq p(\mathbf{x} \bmod \Lambda)$ for all $\mathbf{x} \in \mathbb{R}^L$. The Dirac impulse on the torus $\mathcal{V}$ is defined by the shifting property $\oint_{\mathcal{V}} \delta(\mathbf{w} - \mathbf{z})\overset{\circ}{p}(\mathbf{w})\, d\mathbf{w} = \overset{\circ}{p}(\mathbf{z})$, for any $\mathbf{z} \in \mathcal{V}$ and continuous, $\mathcal{V}$-periodic function $\overset{\circ}{p}$. The following properties of the modΛ operation will be used:

- Distributivity: $\mathbf{x} + (\mathbf{x}' \bmod \Lambda) \bmod \Lambda = \mathbf{x} + \mathbf{x}' \bmod \Lambda$, for all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^L$.
- Uniform noise: if $\mathbf{X}$ is random and uniformly distributed over $\mathcal{V}$, then so is $\mathbf{Z} = \mathbf{X} + \mathbf{Y} \bmod \Lambda$ for any random vector $\mathbf{Y}$.

[1] STDM involves a projection of the host signal onto a lower-dimensional subspace and will be considered in Sec. 8.

4 QIM Watermark Decoding Games

4.1 Lattice decoder

The QIM lattice decoder reduces the data $\mathbf{y}$ to a length-n_B sequence of $\mathcal{V}$-valued statistics:

$$\tilde{\mathbf{y}}(i) \triangleq \alpha \mathbf{y}(i) - \mathbf{d}(i) \mod \Lambda, \quad 1 \le i \le n_B. \tag{4.1}$$

After this reduction step, the dither values $\{\mathbf{d}(i)\}$ are no longer used. Note that further reduction of the data, in the form of a quantization of each $\tilde{\mathbf{y}}(i)$ to the fine lattice Λ_f, would be tantamount to block-level minimum-distance decoding and would generally cause a loss of information about m. Even in the absence of ECC, when blocks are independent and symbols embedded in different blocks may be decoded independently, minimum-distance decoding need not be equivalent to the optimal ML decoding rule [2]. An improved idea would be quantizing each $\tilde{\mathbf{y}}(i)$ to the cosets $\Lambda_0, \cdots, \Lambda_{q-1}$ of the coarse lattice Λ. While this operation is information-lossless in the absence of ECC, it is still information-lossy when an ECC is used.

Denote by p_c the pdf of random vector $\tilde{\mathbf{Y}} \in \mathcal{V}$ given that $c \in \{0, 1, \cdots, q-1\}$ was embedded in the block. Under the hypothesis that message $m \in \mathcal{M}$ was transmitted (and thus $\mathbf{c}(i)$ was embedded in block i for $i = 1, 2, \cdots, n_B$), the random variables $\tilde{\mathbf{Y}}(i)$ are conditionally independent, with respective pdf's $p_{\mathbf{c}(i)}$. To obtain good decoding performance, we need the pdf's $\{p_c, 0 \le c < q\}$ to be "well separated", in a sense made precise below.

4.2 Equivalent Channel Model

Here we derive a model for $\tilde{\mathbf{Y}}$ analogous to the MLAN model of Erez and Zamir [6], in which $\mathbf{W}$ was white Gaussian noise. The outer ECC maps message m into codeword $\mathbf{c}$. In block i, $\tilde{\mathbf{Y}}(i)$ is the sum (modΛ) of the coset vector $\mathbf{z}_{\mathbf{c}(i)}$ and a random vector $\mathbf{V}(i)$:

$$\tilde{\mathbf{Y}}(i) = \mathbf{z}_{\mathbf{c}(i)} + \mathbf{V}(i) \mod \Lambda. \tag{4.2}$$

The random vectors $\mathbf{V}(i)$ are i.i.d. and independent of $\mathbf{c}$. Each vector $\mathbf{V} \in \mathcal{V}$ is the sum (modΛ) of two components:

$$\mathbf{V} = \tilde{\mathbf{E}} + \tilde{\mathbf{W}} \bmod \Lambda, \tag{4.3}$$

where $\tilde{\mathbf{E}}$ is the self-noise due to quantization, which is uniform over the scaled Voronoi cell $(1-\alpha)\mathcal{V}$:

$$p_{\tilde{\mathbf{E}}}(\tilde{\mathbf{e}}) = \frac{1}{(1-\alpha)^L |\mathcal{V}|} 1_{\{\tilde{\mathbf{e}} \in (1-\alpha)\mathcal{V}\}}, \tag{4.4}$$

and $\tilde{\mathbf{W}} \triangleq \alpha \mathbf{W} \bmod \Lambda$ is the scaled attacker's noise, folded into $\mathcal{V}$. Its pdf is

$$p_{\tilde{\mathbf{W}}}(\tilde{\mathbf{w}}) = \sum_{\mathbf{p} \in \Lambda} \frac{1}{\alpha} p_{\mathbf{W}} \left(\frac{\tilde{\mathbf{w}} + \mathbf{p}}{\alpha} \right), \quad \tilde{\mathbf{w}} \in \mathcal{V}. \tag{4.5}$$

[2] The minimum-distance decoding rule coincides with the lattice ML rule in some cases, but not always. As an example where these rules differ, consider scalar QIM with $q = 2$, $\alpha = 1$, and noise pdf $p_W(w) = \frac{1}{2}[\delta(w-a)+\delta(w+a)]$, where $\frac{\Delta}{4} < a < \frac{\Delta}{2}$. The lattice ML decoder is error-free because the rival pdf's $p_0(\tilde{y})$ and $p_1(\tilde{y})$ have disjoint supports. But clearly the minimum-distance decoder is always in error.

The parameter α trades off the amount of self-noise against the attacker's noise at the receiver. When α is small, the self-noise $\tilde{\mathbf{E}}$ dominates the attacker's noise. When $\alpha = 1$, there is no self-noise, and $\mathbf{V} = \tilde{\mathbf{W}}$.

Due to (4.3), (4.4), (4.5), and the independence of $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{W}}$, the pdf of $\mathbf{V}$ is of the form

$$\begin{aligned} p_{\mathbf{V}}(\mathbf{v}) &= \int_{\mathbb{R}^L} \overset{\circ}{p}_{\tilde{\mathbf{E}}}(\mathbf{v} - \alpha\mathbf{w}) p_{\mathbf{W}}(\mathbf{w})\, d\mathbf{w} \\ &= \frac{1}{(1-\alpha)^L |\mathcal{V}|} \oint_{\frac{1}{\alpha}[\mathbf{v}+(1-\alpha)\mathcal{V}]} p_{\mathbf{W}}(\mathbf{w})\, d\mathbf{w}, \quad \mathbf{v} \in \mathcal{V}. \end{aligned} \tag{4.6}$$

From (4.2), we view $\tilde{\mathbf{Y}}(i)$ as the output of a MLAN channel with input $\mathbf{z}_{\mathbf{c}(i)}$ (see Fig. 3). The transition probabilities of the channel are given by

$$p_c(\tilde{\mathbf{y}}) = \overset{\circ}{p}_{\mathbf{V}}(\tilde{\mathbf{y}} - \mathbf{z}_c), \quad 0 \le c < q,\ \tilde{\mathbf{y}} \in \mathcal{V}. \tag{4.7}$$

4.3 Lattice ML decoding rule

We assume that all messages $m \in \mathcal{M}$ are equally likely. Due to the i.i.d. noise model, the decoder is assumed to be able to learn the noise pdf $p_{\mathbf{W}}$ when the number of blocks is large enough. Therefore we shall assume that the decoder knows $p_{\mathbf{W}}$ and implements the lattice ML detection rule, which minimizes the probability of error given $p_{\mathbf{W}}$ and $\tilde{\mathbf{y}}^{1:n_B} \triangleq \tilde{\mathbf{y}}(1), \cdots, \tilde{\mathbf{y}}(n_B)$ [22]:

$$\hat{m} = \operatorname{argmax}_{m \in \mathcal{M}} \underbrace{\prod_{i=1}^{n_B} p_{\mathbf{c}(i)}(\tilde{\mathbf{y}}(i))}_{p(\tilde{\mathbf{y}}(1),\cdots,\tilde{\mathbf{y}}(n_B)|m)}. \tag{4.8}$$

We refer to (4.8) as the lattice ML decoder. Its probability of error is given by

$$P_e = \frac{1}{|\mathcal{M}|} \int_{\mathcal{V}} \cdots \int_{\mathcal{V}} \operatorname{Min}_{m \in \mathcal{M}} \left[\prod_{i=1}^{n_B} p_{\mathbf{c}(i)}(\tilde{\mathbf{y}}(i)) \right] d\tilde{\mathbf{y}}(1) \cdots d\tilde{\mathbf{y}}(n_B) \tag{4.9}$$

where we have used the shorthand $\operatorname{Min}_m h(m) \triangleq \sum_m h(m) - \max_m h(m)$.

4.4 Probability-of-error game

Given the watermarking code, the embedder wants to choose α that minimizes P_e. Similarly, given the code and the optimal α, the attacker wants to choose $p_{\mathbf{W}}$ that maximizes P_e. Equivalently, we may adopt $-\frac{1}{n} \ln P_e$ as the figure of merit of the system. This becomes a maxmin game between the embedder and the attacker:

$$\max_{0 \le \alpha \le 1} \min_{p_{\mathbf{W}}} -\frac{1}{n} \ln P_e(\alpha, p_{\mathbf{W}}), \tag{4.10}$$

where the minimization is subject to the distortion constraint (2.2). Moreover, we shall see in Sec. 4.9 that this minimization can be restricted without loss of optimality to a compact set of pdf's. Existence of a maximum and a minimum is then guaranteed because the cost functional is continuous and bounded, and the maximization and minimization are over compact sets.

For $|\mathcal{M}| > 2$, the union bound may be used to relate P_e to the probabilities of error $P_e(m, m')$ for all $\frac{1}{2}|\mathcal{M}|(|\mathcal{M}| - 1)$ binary tests between messages (m, m'):

$$P_e(\alpha, p_{\mathbf{W}}) \leq \overline{P}_e(\alpha, p_{\mathbf{W}}) \triangleq \frac{1}{|\mathcal{M}|} \sum_{m \neq m'} \frac{1}{2} \int_{\mathcal{V}^{n_B}} \min[p(\tilde{\mathbf{y}}^{1:n_B}|m), p(\tilde{\mathbf{y}}^{1:n_B}|m')] \, d\tilde{\mathbf{y}}^{1:n_B}.$$

The union bound is achieved when $|\mathcal{M}| = 2$.

The special case $n_B = 1$ ($|\mathcal{M}| = q$) corresponds to hard symbol decoding and generally high probability of error – unless L WNR is large. When $|\mathcal{M}| = q = 2$, the cost function in (4.10) becomes

$$-\frac{1}{L} \ln P_e(\alpha, p_{\mathbf{W}}) = -\frac{1}{L} \ln \int_{\mathcal{V}} \frac{1}{2} \min[p_0(\tilde{\mathbf{y}}), p_1(\tilde{\mathbf{y}})] \, d\tilde{\mathbf{y}}. \tag{4.11}$$

4.5 Bhattacharyya game

Even in the simple case $|\mathcal{M}| = 2$, evaluation of the exact probability of error (4.9) requires evaluation of an n-dimensional integral, which unfortunately is infeasible unless n is very small or P_e is relatively large. For large n, a more convenient approach is to use the Bhattacharyya upper bound on P_e [7, 17, 22]. Appendix A presents a succinct derivation of the error bounds stated in this section.

Case $|\mathcal{M}| = q = 2$. For simplicity we begin with the case of two messages: $\mathcal{M} = \{0, 1\}$, $q = 2$, and a repetition outer code. The *per-sample* Bhattacharyya distance between p_0 and p_1 is given by [23]

$$B(p_0, p_1) \triangleq -\frac{1}{L} \ln \int_{\mathcal{V}} \sqrt{p_0(\tilde{\mathbf{y}}) p_1(\tilde{\mathbf{y}})} d\tilde{\mathbf{y}} \tag{4.12}$$

which is zero if $p_0 = p_1$, and positive otherwise. The Bhattacharyya bound

$$P_e \leq P_{e,\text{Bhatt}} \triangleq \frac{1}{2} \, e^{-nB(p_0,p_1)} \tag{4.13}$$

is tight in the exponent when p_0 and p_1 satisfy a certain symmetry property (as will be the case throughout this paper):

$$\lim_{n \to \infty} \left[-\frac{1}{n} \ln P_e \right] = B(p_0, p_1). \tag{4.14}$$

Therefore, the Bhattacharyya bound is frequently used to predict how large n should be to achieve a desired probability of error. [3]

The upper bound in (4.13) is much easier to evaluate than (4.9), because evaluation of $B(p_0, p_1)$ requires only a L-dimensional integration over $\mathcal{V}$, no matter how large $n_B = \frac{n}{L}$ is. Note that in practice, L should be a relatively small number (say 1, 2, or 3); otherwise evaluation of $B(p_0, p_1)$ itself becomes hard. Since $B(p_0, p_1)$ is a function of α and $p_{\mathbf{W}}$, we denote this function as $b(\alpha, p_{\mathbf{W}})$. Instead of (4.10), we solve the simpler (and asymptotically equivalent) problem

$$\max_{0 \leq \alpha \leq 1} \min_{p_{\mathbf{W}}} b(\alpha, p_{\mathbf{W}}), \tag{4.15}$$

[3] A classical example is the binary symmetric channel with crossover probability ϵ. In this case, we have $B(p_0, p_1) = -\ln 2\sqrt{\epsilon(1-\epsilon)}$ and $P_{e,\text{Bhatt}} = (4\epsilon(1-\epsilon))^{n/2}$.

where the minimization is subject to (2.2). The optimal value of $b(\alpha, p_{\mathbf{W}})$ is the *maxmin error exponent* for the P_e game, and we refer to (4.15) as the *Bhattacharyya game*.

Case $|\mathcal{M}| > 2$. The Bhattacharyya bound can be extended to the case of multiple messages. Different bounds can be considered (e.g., random coding bound and expurgated bound [16]—[19]). To simplify the presentation, we fix one that is fairly tractable and provides a useful characterization of decoding performance at very low rates. Our performance metric is the Bhattacharyya parameter for the MLAN channel, which we define as

$$b(\alpha, p_{\mathbf{W}}) \triangleq \frac{1}{q(q-1)} \sum_{c \neq c'} \left[-\frac{1}{L} \ln \int_{\mathcal{V}} \sqrt{p_c(\tilde{\mathbf{y}}) p_{c'}(\tilde{\mathbf{y}})}\, d\tilde{\mathbf{y}} \right] \tag{4.16}$$

i.e., the uniform average of the pairwise Bhattacharyya distances.

Equidistant channels. If $B(p_c, p_{c'})$, the summand of (4.16), is the same for all $c \neq c'$, the channel is said to be *equidistant* [19]. All channels with binary inputs are equidistant. The MLAN channel with the hexagonal lattice of Fig. 2(b) has three inputs and is also equidistant, as can be verified using a simple change of variables in the Bhattacharyya integrals. This holds for any choice of $p_{\mathbf{W}}$.

The Bhattacharyya metric (4.16) is used in conjunction with the normalized minimum distance of the ECC, which is defined as follows. Denote by $d_H(\mathbf{c}, \mathbf{c}')$ the Hamming distance between two codewords $\mathbf{c}$ and $\mathbf{c}' \in \{0, \cdots, q-1\}^{n_B}$ (the number of positions where $\mathbf{c}$ and $\mathbf{c}'$ differ), and

$$\overline{d}_{\min} = \frac{1}{n_B} \min_{\mathbf{c} \neq \mathbf{c}'} d_H(\mathbf{c}, \mathbf{c}') \quad \in [0, 1] \tag{4.17}$$

the normalized minimum-distance of the code. Applying Plotkin's bound, we have $\overline{d}_{\min} \leq \frac{q-1}{q} \frac{|\mathcal{M}|}{|\mathcal{M}|-1}$ for any q-ary code [17, p. 549]. For many low-rate codes, the Plotkin bound is tight. When $|\mathcal{M}| = q$ for instance, any reasonable code satisfies $\overline{d}_{\min} = 1$, i.e., achieves the Plotkin bound with equality. If $|\mathcal{M}|$ tends to infinity as a subexponential function of n, expurgated random codes [16, 17] approach the Plotkin bound, i.e., $\overline{d}_{\min} \to \frac{q-1}{q}$.

Having defined (4.16) and (4.17), we have the upper bound

$$P_e \leq \frac{|\mathcal{M}| - 1}{2} e^{-n\, \overline{d}_{\min}\, b(\alpha, p_{\mathbf{W}})} \tag{4.18}$$

which holds and is tight in the exponent under the aforementioned assumptions of "good" codes with vanishing rate and uniform distribution $Q(c)$ over the alphabet $\{0, \cdots, q-1\}$. For expurgated, zero-rate, random codes, $\frac{q-1}{q} b(\alpha, p_{\mathbf{W}})$ is the expurgated exponent $E_{\text{ex}}(0, Q)$ [17, pp. 156, 540]. For equidistant channels, uniform Q is the optimal choice [19].

4.6 General mathematical formulation of QIM game

The games involving the P_e and Bhattacharyya cost functions P_e in (4.10) and $b(\alpha, p_{\mathbf{W}})$ in (4.16), can be written in the form

$$\max_{0 \leq \alpha \leq 1} \min_{p_{\mathbf{W}}} \beta(\alpha, p_{\mathbf{W}}), \tag{4.19}$$

where the functional β is *convex* in $p_{\mathbf{W}}$ (see Sec. 4.8). In the case of (4.10), we define $\beta = -\frac{1}{n} \ln P_e$. In the case of (4.16), we define $\beta = b$ which is *strictly convex* in $p_{\mathbf{W}}$ (see Sec. 4.8) and is a tight lower bound on $-\frac{1}{n} \ln P_e$.

Since the distortion constraint (2.2) is linear in $p_{\mathbf{W}}$, minimizing $\beta(\alpha, p_{\mathbf{W}})$ over $p_{\mathbf{W}}$ subject to (2.2) is a convex program [24]. The result of this minimization depends on D_1 and D_2 via their ratio $\mathrm{WNR} \triangleq D_1/D_2$. The minimizer will be denoted by

$$\beta^*(\alpha, \mathrm{WNR}) = \min_{p_{\mathbf{W}}} \beta(\alpha, p_{\mathbf{W}}). \tag{4.20}$$

4.7 Ali-Silvey Distances

Consider the P_e cost function specialized to the case $|\mathcal{M}| = q = 2$ in (4.11), and the Bhattacharyya cost function (4.16). Both cost functions are in the class of Ali-Silvey distances [26], also closely related to f-divergences [27, 28]. An f-divergence between two pdf's p_0 and p_1 is a measure of the dispersion of the likelihood ratio. It takes the form $\int p_0(x)\psi\left(\frac{p_1(x)}{p_0(x)}\right) dx$, where ψ is any convex function on $\mathbb{R}^+$. We adopt the notational convention $0\,\psi(\frac{0}{0}) = 0$. In other words, the integral is actually over all x such that $p_0(x)$ and $p_1(x)$ are not simultaneously zero. An Ali-Silvey distance is any functional of the form $C\left(\int p_0\psi\left(\frac{p_1}{p_0}\right)\right)$, where ψ is convex on $\mathbb{R}^+$, and C is an increasing function. We use the equivalent requirement that ψ be concave on $\mathbb{R}^+$ and C be decreasing.

The relation between Ali-Silvey distances and the costs functions (4.11) and (4.16) is as follows.

- The $-\frac{1}{L}\ln P_e$ functional (4.11) is an Ali-Silvey distance with $C(x) = -\frac{1}{L}\ln x$ and $\psi(x) = \frac{1}{2}\min(1, x)$.
- The Bhattacharyya functional (4.16) is an arithmetic average of Ali-Silvey distances, with $C(x) = -\frac{1}{L}\ln x$ and $\psi(x) = \sqrt{x}$.

In the P_e case, notice that $\psi(x)$ is not differentiable at $x = 1$, which presents some technical problems when formulating optimality conditions and using certain convex optimization algorithms. To avoid such problems, one can "round off the corner" at $x = 1$, i.e., replace $\psi(x) = \frac{1}{2}\min(1, x)$ by a differentiable approximation, which can be made arbitrarily accurate in the sup norm.

Denote by $\beta_{\min}$ the minimum value (over all α and all unconstrained pdf's $p_{\mathbf{W}}$) of β. For instance, $\beta_{\min} = -\frac{1}{n}\ln(1 - \frac{1}{|\mathcal{M}|})$ in the P_e case (4.10); this value is achieved when the rival pdf's $p_c, 0 \le c < q$ are identical, eliminating all traces of the watermark at the decoder. When β is the Bhattacharyya parameter (4.16), we have $\beta_{\min} = 0$. For all Ali-Silvey distances, we have $\beta_{\min} = C(\psi(1))$, which is achieved when the rival pdf's are identical.

4.8 Convexity properties

The analysis is facilitated by some important convexity properties of the probability of error and related functionals of $p_{\mathbf{W}}$. These properties are stated below and proved in Appendix B. Also note that $p_{\mathbf{V}}$ in (4.6) is linear in $p_{\mathbf{W}}$.

Lemma 4.1 *(*Convexity properties*):*

(i) $P_e(\alpha, p_{\mathbf{W}})$ *is concave in* $p_{\mathbf{W}}$.

(ii) for $q = 2$, $\exp\{-L\, b(\alpha, p_{\mathbf{W}})\}$ *is concave in* $p_{\mathbf{W}}$.

(iii) $b(\alpha, p_{\mathbf{W}})$ *is convex in* $p_{\mathbf{W}}$.

(iv) $-\frac{1}{n}\ln P_e(\alpha, p_{\mathbf{W}})$ *is convex in* $p_{\mathbf{W}}$.

4.9 Bounded support

Denote by $\overline{\mathcal{V}}$ the topological closure of $\mathcal{V}$. Due to Prop. 4.2 below, from now on we shall restrict our attention to $p_{\mathbf{W}}$ supported over $\frac{1}{\alpha}\overline{\mathcal{V}}$. Choosing a larger domain would introduce excess distortion. The minimum may be achieved by a pdf with impulsive components (on the torus $\frac{1}{\alpha}\mathcal{V}$).

Proposition 4.2 *For any α, there exists $p_{\mathbf{W}}$ minimizing $\beta(\alpha,\cdot)$ in (4.19), such that $p_{\mathbf{W}}(\mathbf{w}) = 0$ for all $\mathbf{w} \notin \frac{1}{\alpha}\overline{\mathcal{V}}$.*

Proof. From (4.3), the noise $\mathbf{W}$ affects the decoder only via the modulo noise $\mathbf{W}' = \mathbf{W} \bmod \frac{1}{\alpha}\Lambda$. Therefore $\beta(\alpha, p_{\mathbf{W}'}) = \beta(\alpha, p_{\mathbf{W}})$. Moreover, $\|\mathbf{w}'\| \le \|\mathbf{w}\|$ for all $\mathbf{w} \in \mathbb{R}^L$, and so if $p_{\mathbf{W}}$ satisfies the distortion constraint (2.2), so does $p_{\mathbf{W}'}$. Hence there is no loss of optimality in searching for a minimizing $p_{\mathbf{W}}$ with bounded support ($\frac{1}{\alpha}\mathcal{V}$), or equivalently, over $\frac{1}{\alpha}\overline{\mathcal{V}}$. The minimization is over all densities defined over $\frac{1}{\alpha}\overline{\mathcal{V}}$. Since the set of all such densities is a compact subset of L_1 and the function β is continuous and bounded from below, a minimizer is guaranteed to exist. □

4.10 Invariance properties

The invariance properties of the code $\mathcal{C}$ and the cost function β play a fundamental role in our analysis. This section defines the basic notions and illustrates them with the examples of Fig. 2.

Recall that an isometry is an L^2 norm preserving linear operator; in finite-dimensional spaces, isometries are unimodular matrices, i.e., matrices with determinant equal to ± 1. Examples include rotation matrices in $\mathbb{R}^L$ and reflections about coordinates axes. A $L \times L$ rotation matrix has $\frac{L(L-1)}{2}$ degrees of freedom. When $L = 2$, the rotation matrix is of the form $\mathsf{G} = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix}$, where the rotation angle $\phi \in [0, 2\pi)$ is the single degree of freedom. The determinant of a rotation matrix is equal to 1. A reflection about the first coordinate axis is represented by the matrix $\mathsf{G} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, whose determinant is equal to -1.

First we examine the invariance properties of the QIM code. Let $\mathcal{G}$ be the set of all isometries in the joint invariance group of the fine and coarse lattices Λ_f and Λ. Clearly, $\mathcal{V}$ and $\mathcal{C}$ inherit the same invariance properties: $\mathsf{G}\mathcal{V} = \mathcal{V}$ and $\mathsf{G}\mathcal{C} = \mathcal{C}$ for all $\mathsf{G} \in \mathcal{G}$. For instance, in Fig. 2(a), $\mathcal{G}$ has *eight* elements: the four rotations by multiples of 90°, as well as the cascade of these rotations with a reflection about the first coordinate axis. In Fig. 2(b), $\mathcal{G}$ has *twelve* elements: the six rotations by multiples of 60°, as well as the cascade of these rotations with a reflection about the first coordinate axis.

For any isometry G and noise pdf $p_{\mathbf{W}}$, define the transformed pdf $p_{\mathsf{G}\mathbf{W}}$ by

$$p_{\mathsf{G}\mathbf{W}}(\mathbf{w}) \triangleq p_{\mathbf{W}}(\mathsf{G}\mathbf{w}), \quad \forall \mathbf{w}. \tag{4.21}$$

Since G is an isometry, $p_{\mathsf{G}\mathbf{W}}$ introduces the same mean-squared distortion as $p_{\mathbf{W}}$. If $p_{\mathbf{W}}$ happens to be invariant to isometries in $\mathcal{G}$ ($p_{\mathsf{G}\mathbf{W}} = p_{\mathbf{W}}$ for all $\mathsf{G} \in \mathcal{G}$), then $p_{\mathbf{W}}$ is said to be $\mathcal{G}$-invariant.

Finally, we turn our attention to the invariance properties of the cost function β. Recall the definitions of β for the P_e and Bhattacharyya games considered in (4.10) and (4.16). By inspection of those expressions, we see that for any $p_{\mathbf{W}}$ ($\mathcal{G}$-invariant or not),

$$\beta(\alpha, p_{\mathsf{G}\mathbf{W}}) = \beta(\alpha, p_{\mathbf{W}}), \quad \forall \mathsf{G} \in \mathcal{G},\ \alpha \in [0,1]. \tag{4.22}$$

The functional $\beta(\alpha, \cdot)$ is thus said to be $\mathcal{G}$-invariant.

Proposition 4.3 [4] *For any α, the minimum in the β-game (4.19) is achieved by some $\mathcal{G}$-invariant $p_{\mathbf{W}}$. Moreover, no other solution is possible if β is strictly convex.*

Proof. The claim is a consequence of the $\mathcal{G}$-invariance and convexity properties of β. Since the self-noise is uniform over $\mathcal{V}$, its pdf $p_{\tilde{\mathbf{E}}}$ is $\mathcal{G}$-invariant. It therefore follows from (4.3) that $p_{\mathbf{V}}$ is $\mathcal{G}$-invariant as well. Now define the isometry-averaged pdf

$$\overline{p}_{\mathbf{W}}(\mathbf{w}) = \frac{1}{|\mathcal{G}|} \sum_{\mathsf{G} \in \mathcal{G}} p_{\mathsf{G}\mathbf{W}}(\mathbf{w})$$

which is $\mathcal{G}$-invariant. By linearity of the distortion functional in (2.2), $\overline{p}_{\mathbf{W}}$ also satisfies the distortion constraint (2.2). Exploiting successively the convexity and the $\mathcal{G}$-invariance of $\beta(\alpha, \cdot)$, we may write

$$\begin{aligned} \beta(\alpha, \overline{p}_{\mathbf{W}}) &\leq \frac{1}{|\mathcal{G}|} \sum_{\mathsf{G} \in \mathcal{G}} \beta(\alpha, p_{\mathsf{G}\mathbf{W}}) \\ &= \beta(\alpha, p_{\mathbf{W}}), \end{aligned}$$

i.e., $\overline{p}_{\mathbf{W}}$ is at least as bad as $p_{\mathbf{W}}$. If β is strictly convex, equality is achieved if and only if $p_{\mathbf{W}} = \overline{p}_{\mathbf{W}}$. □

5 Scalar QIM, Memoryless Attacks

We first consider the simplest nontrivial problem: scalar QIM with binary input alphabet ($q = 2$) and block length $L = 1$ (therefore $n_B = n$). This design maximizes the Euclidean distance between the cosets Λ_0 and Λ_1 [29]. Denote by Δ the quantizer step size for the coarse lattice Λ; its Voronoi cell is $\mathcal{V} = [-\frac{\Delta}{2}, \frac{\Delta}{2})$. The distance between Λ_0 and Λ_1 is $|z_1 - z_0| = \frac{\Delta}{2}$; without loss of generality, we take $z_0 = -z_1 = \frac{\Delta}{4}$. The embedding distortion (3.3) is $D_1 = \frac{\Delta^2}{12}$. The distortion constraint (2.2) simplifies to

$$\int_{\mathbb{R}} w^2 p_W(w)\, dw \leq D_2. \tag{5.1}$$

The noise in the MLAN channel is given by

$$V = \tilde{E} + \tilde{W} \bmod \Delta \quad \in \left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right). \tag{5.2}$$

Due to Prop. 4.2, we can restrict our attention to p_W with support $[-\frac{\Delta}{2\alpha}, \frac{\Delta}{2\alpha}]$, in which case (4.6) becomes

$$p_V(v) = \frac{1}{(1-\alpha)\Delta} \oint_{\frac{v}{\alpha} - \frac{(1-\alpha)\Delta}{2\alpha}}^{\frac{v}{\alpha} + \frac{(1-\alpha)\Delta}{2\alpha}} p_W(w)\, dw, \quad |v| \leq \frac{\Delta}{2}. \tag{5.3}$$

The rival pdf's for the decoder are given by

$$p_0(\tilde{y}) = \mathring{p}_V\left(\tilde{y} + \frac{\Delta}{4}\right), \quad p_1(\tilde{y}) = \mathring{p}_V\left(\tilde{y} - \frac{\Delta}{4}\right), \quad \tilde{y} \in \left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right). \tag{5.4}$$

Fig. 4 shows p_0 and p_1 for a Gaussian attack and watermark-to-noise ratio WNR = 0.1.

[4] Our original statement of this proposition only involved invariance with respect to rotations. This stronger statement is due to Ton Kalker.

5.1 Probability-of-error game

Let β be either the P_e cost functional (4.10) or a negative f-divergence. The minmax game (4.19) between the embedder and the attacker takes the form

$$\min_{0\le\alpha\le 1}\ \max_{p_W} \beta(\alpha, p_W) \tag{5.5}$$

where the minimization is subject to (5.1). Our first result is a special case of Prop. 4.3.

Proposition 5.1 *For any α, the minimum in (5.5) is achieved by some p_W that is symmetric around 0. Moreover, no other solution is possible if β is strictly convex.*

If p_W are symmetric around 0, so is p_V in (5.2), and the rival pdf's $p_0(\tilde{y})$ and $p_1(\tilde{y})$ have means $z_0 = -\frac{\Delta}{4}$ and $z_1 = \frac{\Delta}{4}$, respectively.

The next result is about the maximization over α, and is useful only for $\mathrm{WNR} \le \frac{4}{3}$.

Proposition 5.2 *The maximizing α in (5.5) is smaller than $\overline{\alpha} = \sqrt{\frac{3}{4}\mathrm{WNR}}$. For any $\alpha \ge \overline{\alpha}$, we have $\min_{p_W} \beta(\alpha, p_W) = \beta_{\min}$.*

Proof. An ideal p_W for the attacker would be one that assigns mass $\frac{1}{2}$ to $w = \frac{\Delta}{4\alpha}$ and to $w = -\frac{\Delta}{4\alpha}$, because p_0 and p_1 are identical in this case, and therefore $\beta(\alpha, p_W) = \beta_{\min}$, which is the worst possible value. To be feasible, such p_W must satisfy the distortion constraint $\mathbb{E}(W^2) = (\frac{\Delta}{4\alpha})^2 = \frac{3D_1}{4\alpha^2} \le D_2$. This is possible if and only if $\alpha \ge \sqrt{\frac{3D_1}{4D_2}} = \overline{\alpha}$. □

5.2 Negative f-divergence game

For scalar QIM with $q = 2$, the P_e cost function (with with $n_B = 1$) in (4.11) and the Bhattacharyya cost function in (4.16) are negative f-divergences:

$$\beta(\alpha, p_W) = -\ln \int_{-\Delta/2}^{\Delta/2} \mathring{p}_V(\tilde{y} - \Delta/4)\, \psi\left(\frac{\mathring{p}_V(\tilde{y} + \Delta/4)}{\mathring{p}_V(\tilde{y} - \Delta/4)}\right) d\tilde{y}. \tag{5.6}$$

The optimal p_W may be derived analytically when $\alpha = 1$, as stated in Prop. 5.3 below. The proof may be found in Appendix C.

Proposition 5.3 *When $\alpha = 1$, the maximizing p_W allocates mass to at most four values, $w \in \{\pm w^*, \pm(\frac{\Delta}{2} - w^*)\}$:*

$$p_W(w) = \frac{1-a}{2}\left[\delta(w - w^*) + \delta(w + w^*)\right] + \frac{a}{2}\left[\delta\left(w - \frac{\Delta}{2} + w^*\right) + \delta\left(w + \frac{\Delta}{2} - w^*\right)\right] \tag{5.7}$$

where $0 \le w^* \le \frac{\Delta}{4}$ and $0 \le a \le \frac{1}{2}$. The value of the parameters w^* and a depends on WNR.

(i) If $\mathrm{WNR} \le \frac{4}{3}$, *the minimum of $\beta(1, \cdot)$ is achieved by*

$$w^* = \frac{\Delta}{4}, \quad a = 0 \quad \Rightarrow \quad \beta^*(1, \mathrm{WNR}) = -\ln \psi(1) = \beta_{\min}.$$

(ii) If $\mathrm{WNR} > \frac{4}{3}$*, then*

$$\begin{aligned} w^* &= \frac{\Delta}{4}\left\{1-\sqrt{1-\frac{4}{3\,\mathrm{WNR}}}\right\} \quad \sim \frac{\Delta}{6\,\mathrm{WNR}} \quad \text{as } \mathrm{WNR}\to\infty, \\ a &= \frac{1}{2}\left\{1-\sqrt{1-\frac{4}{3\,\mathrm{WNR}}}\right\} \quad \sim \frac{1}{3\,\mathrm{WNR}} \quad \text{as } \mathrm{WNR}\to\infty. \end{aligned}$$

For the Bhattacharyya game,

$$\beta^*(1,\mathrm{WNR}) = \frac{1}{2}\ln\frac{3\,\mathrm{WNR}}{4}.$$

For the P_e *game with* $n_B = 1$*,*

$$\begin{aligned} \beta^*(1,\mathrm{WNR}) = -\ln P_e &= -\ln\frac{1}{2}\left\{1-\sqrt{1-\frac{4}{3\,\mathrm{WNR}}}\right\} \\ &\downarrow \quad \ln(3\,\mathrm{WNR}) \quad \text{as } \mathrm{WNR}\to\infty. \end{aligned}$$

Proof: see Appendix C.

Fig. 5 shows the rival $p_0(\tilde{y})$ and $p_1(\tilde{y})$ that result from the optimal "four-delta attack" in (5.7). Case (i) of Prop. 5.3 is related to Prop. 5.2: the attacker does not need to use all of his distortion budget when $\mathrm{WNR} < \frac{4}{3}$, because the devastating "two-delta" attack with impulses at $\pm\frac{\Delta}{4}$ only introduces distortion $\mathbb{E}(W^2) = \frac{3}{4}D_1$. We conclude that QIM without distortion compensation [1] completely fails against an intelligent attacker with power $D_2 \geq \frac{3}{4}D_1$; this holds whether or not the embedder uses an outer ECC.

At the other extreme, when $\mathrm{WNR}\to\infty$, we have $w^*\to 0$ and $a \sim \frac{1}{3\,\mathrm{WNR}}$. The optimal p_W could be approximated with the "three-delta" attack that would result from selecting $w^* = 0$ and $a = \frac{1}{3\,\mathrm{WNR}}$ [5]. Note the continuous evolution of the optimal p_W as WNR is increased from $\frac{4}{3}$ to infinity.

In the P_e case ($n_B = 1$), the error probability obtained in (ii) is almost one fourth that obtained in [13] under different assumptions: minimum-distance detector instead of ML detector, and attacker's pdf constrained to a "three-delta" class with mass $1-a$ at $w=0$ and $\frac{a}{2}$ at $\pm\Delta/4$ [6].

5.3 Numerical Results

We have solved the Bhattacharyya game numerically using the convex programming resources of TOMLAB [30]. Integrals over $[-\frac{\Delta}{2},\frac{\Delta}{2}]$ are approximated with finite sums: we used a uniform discretization of the interval with 43 points. Fig. 6(a) compares the Bhattacharyya bound with the actual probability of error (for $n = 15$) computed via Monte-Carlo simulations. The Bhattacharyya

[5] The corresponding Bhattacharyya distance is $\frac{1}{2}\ln\frac{9\,\mathrm{WNR}}{4(3\,\mathrm{WNR}-1)}$.

[6] The Bhattacharyya parameter for that particular attack is $\ln\frac{3\,\mathrm{WNR}}{4}$, i.e., twice the minimizing value in (ii). It is also instructive to look at the effects of that attack on a length-n sequence. For that attack, the ML detector can err only when *all* n noise samples have absolute value $\Delta/4$; the probability of this event is equal to a^n (and the value of $\|\mathbf{w}\|^2$ is atypically large). As discussed in the text, when the ML decoder is used and WNR is high, the optimal attack puts mass in the vicinity of $\pm\Delta/2$.

bound differs from the actual probability of error only by approximately 0.7 in $\log_{10}$ units [7]. Fig. 6(b) shows the optimal α as a function of WNR, and Fig. 7 shows the worst-case p_W for three values of WNR. Note that the optimal α is neither $\frac{\mathrm{WNR}}{\mathrm{WNR}+1}$ [2, 10] nor $\sqrt{\frac{\mathrm{WNR}}{\mathrm{WNR}+2.7}}$ [5], both of which are the results of optimization for coding problems with i.i.d. Gaussian W. An important observation is that the worst-case p_W is strongly non-Gaussian in all examples of Fig. 7.

5.4 Impulsive noise

While closed-form expressions for the worst p_W are not available for $\alpha < 1$, useful insights are obtained by analyzing the performance of impulsive noise, specifically the following 3-delta attack pdf:

$$p_W^{(\alpha)} = (1-a)\delta(w) + \frac{a}{2}\left[\delta\left(w - \frac{\Delta}{2\alpha}\right) + \delta\left(w + \frac{\Delta}{2\alpha}\right)\right] \tag{5.8}$$

with $a = \frac{\alpha^2}{3\mathrm{WNR}}$. This attack is feasible for all $\mathrm{WNR} \geq \frac{4}{3}$. From Prop 5.3, we also know that this attack is asymptotically optimal for large WNR, when $\alpha = 1$. The corresponding Bhattacharyya distance is $\beta^*(1, \frac{\mathrm{WNR}}{\alpha^2}) = \frac{1}{2}\ln\frac{3\mathrm{WNR}}{4\alpha^2}$. However we would also like to know whether the embedder can improve performance by choosing a smaller value of α. Below we derive simple expressions for the Bhattacharyya distance under the 3-delta attack, valid for any α. These expressions provide upper bounds on the value of the Bhattacharyya game, and they are tight for large WNR. The general idea, used to prove Prop. 5.4 below, is to write

$$\max_{0\leq\alpha\leq 1}\min_{p_W} b(\alpha, p_W) \leq \max_{0\leq\alpha\leq 1} b(\alpha, p_W^{(\alpha)}). \tag{5.9}$$

As the proof of Prop. 5.4 shows (see Appendix D), the resulting p_V is piecewise-constant, and closed-form expressions for the Bhattacharyya distances can be derived.

Proposition 5.4 *The value of the Bhattacharyya game ($\psi(x) = \sqrt{x}$) is upper-bounded by*

$$\max_{0\leq\alpha\leq 1}\min_{p_W} b(\alpha, p_W) \leq \frac{1}{2}\ln(3\,\mathrm{WNR}) - \frac{1}{2}\ln\left(1 - \frac{1}{12\mathrm{WNR}}\right), \quad \forall\,\mathrm{WNR} \geq \frac{4}{3}. \tag{5.10}$$

The bound (5.10) is quite tight. For instance, when $\mathrm{WNR} = \frac{4}{3}$, the bound is approximately equal to 0.379. A numerical evaluation of the left side of (5.10) yields $\max_{0\leq\alpha\leq 1}\min_{p_W} b(\alpha, p_W) \approx 0.326$. Furthermore, let us compare (5.10) with the case of Gaussian p_W, for which the optimal α tends to 1 at high WNR:

$$b(1, p_W) \sim \frac{3\,\mathrm{WNR}}{8} \gg \frac{1}{2}\ln\left(\frac{3\,\mathrm{WNR}}{4}\right) \sim \min_{p_W} b(1, p_W) \quad \text{as } \mathrm{WNR} \to \infty.$$

Thus the Bhattacharyya distance increases *linearly* with WNR under Gaussian p_W but only *logarithmically* under the impulsive-noise attack.

To summarize the main practical results in this section:

[7] Also shown in Fig. 6(a) is an estimate of P_e calculated using a Gaussian approximation to V (with the same variance as the actual V). While such "estimates" are often used, they have no theoretical justification, especially in the large-deviations regime (large n). In Fig. 6(a), $n = 15$, and the Gaussian "estimate" is incorrect by four orders of magnitude.

- Impulsive-noise attacks are far more damaging than Gaussian attacks at moderate-to-high WNR's.
- The choice $\alpha = 1$ studied in the original QIM paper [1] performs catastrophically when WNR $\leq \frac{4}{3}$.

6 Attacks with Memory

In this section we return to our general setup with L-dimensional lattices, and allow the noise L-vector $\mathbf{W}$ to have dependent components (i.e., memory). We wish to quantify the potential benefits of this strategy from the attacker's perspective. Recall from Sec. 4.10 that the QIM code is invariant to isometries in a set $\mathcal{G}$, and that for any $\alpha \in [0,1]$, the maximum of $\beta(\alpha,\cdot)$ is achieved by some $\mathcal{G}$-invariant $p_{\mathbf{W}}$ (Prop. 4.3).

We have evaluated the Bhattacharyya distance as a function of α for the scalar QIM code ($L=1, q=2$), the cubic QIM code of Fig. 2(a) ($L=2, q=2$), and the hexagonal code of Fig. 2(b) ($L=2, q=3$). Again the optimization was implemented using TOMLAB; in the hexagonal QIM case, a hexagonal grid with 43 points along each diameter (total of 1331 grid points) was used to discretize $\mathcal{V}$ and perform circular convolutions. The results are summarized in Tables 1 and 2 for WNR = 1. The first table also includes results for STDM, which will be discussed in Sec. 8.

The minmax optimal $p_{\mathbf{W}}$ is shown in Fig. 8(a)(b) for the cubic and hexagonal QIM schemes considered. For cubic QIM, this $p_{\mathbf{W}}$ is concentrated along the two main diagonals in the $\mathbf{W}$ plane. This $p_{\mathbf{W}}$ is not memoryless; the Bhattacharyya distance is 0.195, i.e., a drop from 0.214 obtained under the worst *memoryless* attack of Sec. 5. For hexagonal QIM, as might be expected, the worst $p_{\mathbf{W}}$ exhibits six preferred directions.

To summarize this section, memory helps the attacker to develop efficient surgical attacks against the lattice code by selecting the most damaging directions for his noise vector.

7 Lattice QIM with Randomized Rotation

Finally, we provide the watermark embedder with a new strategy (besides dithering) to improve robustness against attacks with memory. Namely, we allow a randomized rotation of the lattice in addition to dithering. The issue is now to evaluate the usefulness of such randomization.

The basic QIM scheme with randomized rotation is diagrammed in Fig. 9. For each block, a different G is generated from the uniform distribution on the set $\mathcal{G}_L$ of *all* $L \times L$ rotation matrices (a continuum, unlike the discrete invariant set $\mathcal{G}$ considered in the previous sections). The secret θ in Fig. 1 consists thus of the pair $(\mathbf{d}, \mathsf{G})$. Consider for instance the cubic lattice with $q=2$ and antipodal dither vectors $\mathbf{z}_0 = -\mathbf{z}_1$. In Sec. 6, G is deterministic, and so the vector $\mathbf{z}_0 = (-\frac{\Delta}{4}, -\frac{\Delta}{4}, \ldots, -\frac{\Delta}{4})$ is known to the attacker. In contrast, when G is randomized, the vector $\mathbf{z}_0$ is generated from the uniform pdf over the L-sphere with radius $\sqrt{L}\frac{\Delta}{4}$.

7.1 General Lattice QIM

For arbitrary L-dimensional lattices and arbitrary q, we now show that the worst attack pdf is isotropic. Denote by $P_e(\alpha, p_{\mathbf{W}}|\mathsf{G}^{1:n_B})$ the probability of error of the QIM scheme conditioned on a

particular realization $\mathsf{G}^{1:n_B}$ of the n_B rotation matrices, and by $P_e(\alpha, p_{\mathbf{W}})$ the probability of error averaged over all $\mathsf{G}^{1:n_B}$. Also denote by $\beta(\alpha, p_{\mathbf{W}}|\mathsf{G})$ the Bhattacharyya cost function (4.16) under a fixed value of G, and by

$$\beta(\alpha, p_{\mathbf{W}}) = \int_{\mathcal{G}_L} \beta(\alpha, p_{\mathbf{W}}|\mathsf{G})\, d\mu(\mathsf{G}) \tag{7.1}$$

its average with respect to the uniform distribution on $\mathcal{G}_L$. As shown in see Appendix A, the bound (4.18) on $P_e(\alpha, p_{\mathbf{W}})$ still holds.

Proposition 7.1 *Under uniform randomization of the lattice rotation, the minimum over $p_{\mathbf{W}}$ in the $-\frac{1}{n}\ln P_e(\alpha, p_{\mathbf{W}})$ and $\beta(\alpha, p_{\mathbf{W}})$ games is achieved by an isotropic pdf.*

Proof. The Jacobian of G is unity, and therefore the functional $\beta(\alpha, \cdot)$ in (7.1) is $\mathcal{G}_L$-invariant:

$$\beta(\alpha, p_{\mathbf{W}}) = \beta(\alpha, p_{\mathsf{G}\mathbf{W}}), \quad \forall \mathsf{G} \in \mathcal{G}_L. \tag{7.2}$$

For each value of G, the functional $\beta(\alpha, \cdot|\mathsf{G})$ is convex (by Lemma 4.1), and therefore $\beta(\alpha, \cdot)$ in (7.1) is convex as well. To complete the proof, we apply Prop. 4.3. Since $\beta(\alpha, \cdot)$ is $\mathcal{G}_L$-invariant and convex, its minimum is achieved by a $\mathcal{G}_L$-invariant (i.e., isotropic) $p_{\mathbf{W}}$. The same arguments apply to the cost function $-\frac{1}{n}\ln P_e(\alpha, p_{\mathbf{W}})$. □

From Prop. 7.1, we conclude that the worst noise pdf $p_{\mathbf{W}}$ is characterized by a pdf $p_R(\rho), \rho \geq 0$, in the radial direction: $p_{\mathbf{W}}(\mathbf{w}) = p_R(\|\mathbf{w}\|)$. Hence, with a little abuse of notation, we write the optimization problem (4.19) as

$$\max_{0\leq\alpha\leq 1} \min_{p_R} \beta(\alpha, p_R) \tag{7.3}$$

where $\beta(\alpha, p_R) \triangleq \int_{\mathcal{G}_L} \beta(\alpha, p_R|\mathsf{G})\, d\mu(\mathsf{G})$. The minimization over p_R is subject to the distortion constraint

$$\frac{1}{L}\int_0^\infty \rho^2 p_R(\rho)\, d\rho \leq D_2. \tag{7.4}$$

Due to (4.6), the mapping from p_R to $p_{\mathbf{V}}$ is linear:

$$p_{\mathbf{V}}(\mathbf{v}) = \int_0^\infty p_R(\rho) p_{\mathbf{V}|R}(\mathbf{v}|\rho)\, d\rho, \quad \mathbf{v} \in \mathcal{V} \tag{7.5}$$

where

$$p_{\mathbf{V}|R}(\mathbf{v}|\rho) = \mathbb{E}_{\mathbf{W}|\rho}\left[\mathring{p}_{\tilde{\mathbf{E}}}(\mathbf{v} - \alpha\mathbf{W})\right] = \frac{1}{|S_L(\rho)|}\int_{S_L(\rho)} \mathring{p}_{\tilde{\mathbf{E}}}(\mathbf{v} - \alpha\mathbf{w})\, d\mathbf{w} \tag{7.6}$$

is the integration kernel, and $S_L(\rho)$ denotes the centered L-dimensional sphere with radius ρ.

Note that Prop. 4.2 does not apply here because the attacker does not know the lattice orientation. The optimal radial pdf $p_R(\rho)$ has unbounded support in general.

Lemma 7.2 *The P_e and Bhattacharyya functionals $\beta(\alpha, p_R)$ are convex in p_R.*

Proof: analogous to the proof of Lemma 4.1(i) and (ii), exploiting the linearity of the mapping (7.5) from p_R to $p_{\mathbf{V}}$. □

7.2 Numerical examples

Again we solve the Bhattacharyya game ($\psi(x) = \sqrt{x}$) using the convex programming resources of [30]. Consider the cubic QIM problem with $q = 2$. In the case of a code with deterministic rotation, recall from Fig. 8(a) that the optimal directions for the noise vector $\mathbf{w}$ are diagonal. In the case of a code with randomized rotations, the worst attack pdf is spread uniformly in all directions; its radial pdf is depicted in Fig. 10. The value of the Bhattacharyya game is 0.235, compared with only 0.195 when no lattice randomization was used (Sec. 6). The benefits of randomized rotations are thus clear, as is the non-Gaussian nature of the worst $p_{\mathbf{W}}$ (p_R would be Rayleigh in this case).

8 Dithered STDM

STDM is a projection method which first forms the dot product of $\mathbf{s}$, a length-L block of host data, with $\mathbf{u}$, an arbitrary unit-norm vector. This is followed by application of scalar QIM to the dot product $\mathbf{s} \cdot \mathbf{u}$. Chen and Wornell [2] used $q = 2$, $\alpha = 1$, and no external dither $\mathbf{d}$. The code rate $R \leq \frac{1}{L}$ is typically low.

We choose $\mathbf{u}$ uniformly distributed on the unit sphere, analogously to the randomized lattice rotation method in the previous section. Note that while convenient and nearly optimal for large block length L, the choice $\alpha = 1$ is not optimal. We shall therefore not impose the restriction $\alpha = 1$ here. As we shall see, the STDM problem is insightful and mathematically tractable, even for large L.

8.1 Worst Attack

The STDM quantizer step size is $\Delta = \sqrt{12LD_1}$. For our statistical performance analysis, without loss of generality we may assume that quantization is applied to the first component s_1 of the host signal, i.e., $\mathbf{u} = (1, 0, 0, \cdots, 0)$. The results of Sec. 7 extend to the STDM case as follows [11].

The worst noise pdf $p_{\mathbf{W}}$ is isotropic, with radial pdf p_R. We have $P_e(\alpha, p_R) \leq \exp\{-n\beta(\alpha, p_R)\}$, where

$$\beta(\alpha, p_R) = -\frac{1}{L} \ln \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \mathring{p}_V(v - \Delta/4)\, \psi\left(\frac{\mathring{p}_V(v + \Delta/4)}{\mathring{p}_V(v - \Delta/4)}\right) dv \tag{8.1}$$

(with equality if $n = L$ and $\psi(x) = \frac{1}{2}\min(1, x)$). Similarly to (7.5), we have $p_V(v) = \int_0^\infty p_R(\rho) p_{V|R}(v|\rho)\, d\rho$ for $-\frac{\Delta}{2} \leq v < \frac{\Delta}{2}$, where the integration kernel is given by a simple 1-D integral formula (instead of the L-dimensional integral in (7.6)):

$$p_{V|R}(v|\rho) = \mathbb{E}_{W_1|\rho}\,[\mathring{p}_{\tilde{E}}(v - \alpha W_1)] = \int_{-\rho}^{\rho} \mathring{p}_{\tilde{E}}(v - \alpha w_1) \frac{c_L}{\rho} \left(1 - \frac{w_1^2}{\rho^2}\right)^{\frac{L-3}{2}} dw_1, \quad |v| \leq \frac{\Delta}{2} \tag{8.2}$$

and $c_L = 2^{2-L}\Gamma(L-1)/\Gamma^2(\frac{L-1}{2})$ (hence $c_2 = \frac{1}{\pi}, c_3 = \frac{1}{2}$, and, by Stirling's formula, $c_L \sim \sqrt{2\pi(L-1)}$ as $L \to \infty$). Furthermore, the expression (8.2) for the kernel $p_{V|R}$ reduces to a finite sum with up to $\lceil 2\rho/\Delta \rceil$ terms when $\alpha = 1$ ($p_{\tilde{E}}$ is the Dirac impulse).

8.2 AWGN + Delta Attack

For dithered STDM, if $p_{\mathbf{W}}$ is AWGN and $\alpha = 1$, we have, in the limit of large L,

$$b(1, p_R) \sim \frac{1}{L} \frac{(\Delta/2)^2}{8D_2} = \frac{3\,\mathrm{WNR}}{8}. \tag{8.3}$$

In general, finding the worst p_R analytically is difficult, even in the STDM case. Nevertheless, in this section we show that for large L, an effective choice for $\mathbf{W}$ is an "AWGN+Delta" distribution: the mixture of a mass distribution at zero and an AWGN with large variance. Such attacks have been been used in the spread-spectrum communications literature and are very effective against block codes – see Viterbi's classical paper [32], for instance.

Let $\beta_G(\alpha, \mathrm{WNR})$ denote the value of $\beta(\alpha, p_W)$ for the scalar QIM problem of Sec. 5 when p_W is Gaussian $\mathcal{N}(0, D_2)$ and $\mathrm{WNR} = \frac{\Delta^2/12}{D_2}$. Define $c^* = \max_{0\le\alpha\le1} \beta_G(\alpha, 1)$ and let α^* achieve the maximum above. It can be verified numerically that $\alpha^* \approx \frac{1}{2}$ and $c^* \approx 0.2435$.

Lemma 8.1 *For all $\alpha \in (0, 1]$ and $\mathrm{WNR} \ge 1/L$, there exists a feasible AWGN+Delta distribution $p_{\mathbf{W}}$ such that*

$$-\frac{1}{L} \ln P_e(\alpha, p_{\mathbf{W}}) \sim \beta(\alpha, p_{\mathbf{W}}) \le \frac{1}{L}[\beta_G(\alpha, 1) + \ln(L\,\mathrm{WNR})].$$

The weight of the AWGN component is $\epsilon = \frac{1}{L\,\mathrm{WNR}}$ and its variance is $D = D_2/\epsilon$.

Proof. Let $D \ge D_2$ be a free parameter and $\tilde{p}_{\mathbf{W}} = \mathcal{N}(0, DI_L)$. Any AWGN+Delta distribution on the boundary of the feasible set takes the form $p_{\mathbf{W}}(\mathbf{w}) = (1-\epsilon)\delta(\mathbf{w}) + \epsilon\tilde{p}_{\mathbf{W}}(\mathbf{w})$ where $\epsilon D = D_2$ to satisfy the distortion constraint (2.2) with equality. Then

$$\begin{aligned} e^{-L\beta(\alpha, p_{\mathbf{W}})} &\ge (1-\epsilon)\, e^{-L\beta(\alpha,\delta)} + \epsilon\, e^{-L\beta(\alpha,\tilde{p}_{\mathbf{W}})} \\ &\ge \epsilon\, e^{-L\beta(\alpha,\tilde{p}_{\mathbf{W}})} \end{aligned} \tag{8.4}$$

where the first inequality is due to the concavity of the functional $e^{-L\beta(\alpha,\cdot)}$ (from Lemma 4.1(ii)). Thus (8.4) yields

$$\beta(\alpha, p_{\mathbf{W}}) \le \frac{1}{L} \ln \epsilon^{-1} + \beta(\alpha, \tilde{p}_{\mathbf{W}}).$$

The quantizer step size for STDM is $\Delta = \sqrt{12LD_1}$, and therefore $\beta(\alpha, \tilde{p}_{\mathbf{W}}) = \frac{1}{L}\beta_G(\alpha, \epsilon L\,\mathrm{WNR})$. Choosing $\epsilon = \frac{1}{L\,\mathrm{WNR}}$ proves the claim. □

When $\psi(x) = \sqrt{x}$, we obtain the following proposition.

Proposition 8.2 *The error exponent for the STDM Bhattacharyya game satisfies*

$$\max_{0\le\alpha\le1} \min_{p_R} b(\alpha, p_R) \le \frac{1}{L}[c^* + \ln(L\,\mathrm{WNR})], \quad \forall L \ge \frac{1}{\mathrm{WNR}} \tag{8.5}$$

where $c^* \approx 0.2435$. The upper bound is achieved by the AWGN+Delta distribution of Lemma 8.1.

The upper bound tends to zero for large L, indicating that the AWGN + Delta attack is a powerful surgical attack against STDM – much more so than the AWGN attack whose performance was given by (8.3). We have obtained a similar bound for general lattice QIM with randomized rotation; derivations and results are not reported here due to space constraints. Therefore, even randomization of the lattice orientation does not suffice to guarantee good performance against an intelligent adversary. A possible improved strategy for the embedder is mentioned in Sec. 10.

9 Capacity

In this section, we apply our analytical framework to the problem of finding the noise pdf $p_{\mathbf{W}}$ that minimizes capacity of a given lattice QIM system (as opposed to the probability of error studied so far); the minmax value of α can also be derived in this framework.

Referring to the MLAN channel of Fig. 3, the maximum rate of reliable transmission between encoder and decoder is given by $R = I(\mathbf{Z}; \tilde{\mathbf{Y}})$ where $\mathbf{Z}, \tilde{\mathbf{Y}} \in \mathcal{V}$, and

$$I(\mathbf{Z}; \tilde{\mathbf{Y}}) = \int_{\mathcal{V}} \int_{\mathcal{V}} p_{\mathbf{Z}}(\mathbf{z}) \mathring{p}_{\mathbf{V}}(\mathbf{z} - \tilde{\mathbf{y}}) \ln \frac{\mathring{p}_{\mathbf{V}}(\mathbf{z} - \tilde{\mathbf{y}})}{p_{\mathbf{Z}}(\mathbf{z})} \, d\tilde{\mathbf{y}} d\mathbf{z} \tag{9.1}$$

denotes mutual information for the MLAN channel with input distribution $p_{\mathbf{Z}}$ and channel law $p_{\mathbf{V}}$, which depends on α and $p_{\mathbf{W}}$. The problem reduces to a mutual-information game:

$$\max_{0 \le \alpha \le 1} \max_{p_{\mathbf{Z}}} \min_{p_{\mathbf{W}}} I(\mathbf{Z}; \tilde{\mathbf{Y}}) \tag{9.2}$$

where $p_{\mathbf{V}}$ relates linearly to $p_{\mathbf{W}}$ via (4.6). The cost function is convex in $p_{\mathbf{W}}$. The minimization over $p_{\mathbf{W}}$ is subject to the distortion constraint (2.2), as well as to possible memory constraints.

If the alphabet for $\mathbf{Z}$ is the entire Voronoi region $\mathcal{V}$ (as opposed to some discrete subset $\{\mathbf{z}_1, \cdots, \mathbf{z}_q\}$ as used so far), some simplifications arise. Observe that [6]

$$I(\mathbf{Z}; \tilde{\mathbf{Y}}) = h(\tilde{\mathbf{Y}}) - h(\tilde{\mathbf{Y}}|\mathbf{Z}) = h(\tilde{\mathbf{Y}}) - h(\mathbf{V}) \le \ln |\mathcal{V}| - h(\mathbf{V}) \tag{9.3}$$

where $h(X) = -\int p_X(x) \ln p_X(x) \, dx$ denotes the differential entropy of a random variable X. Equality holds in (9.3) if $\mathbf{Z}$ is uniform over $\mathcal{V}$ (then $\tilde{\mathbf{Y}}$ is also uniform over $\mathcal{V}$, due to the uniform noise property in Sec. 3). Therefore the maximum in (9.2) is achieved by uniform $p_{\mathbf{Z}}$, and (9.2) can be reduced to our "standard game" (4.19), in which the cost function $\beta(\alpha, p_{\mathbf{W}})$ is the differential entropy $h(\mathbf{V})$, which is concave in $p_{\mathbf{W}}$. The relevant results of the previous sections apply to this cost function as well, including symmetry properties of the worst noise pdf, isotropy under randomized lattice rotations, and numerical optimization methods.

We conclude this section by bridging some results from [10] and [6]. In [10], the worst *memoryless* noise against *any* coding scheme (not necessarily lattice QIM) under squared–error distortion constraints was shown to be Gaussian. Likewise, the worst *blockwise-memoryless* noise is i.i.d. Gaussian. On the other hand, Erez and Zamir [6] were concerned solely about AWGN and proved the existence of a sequence of lattice QIM codes with increasing dimension that achieve the unconstrained capacity $C = \frac{1}{2} \ln(1 + D_1/D_2)$. Proposition 9.1 below shows that the attacker gains no advantage by using a non-Gaussian distribution against the Erez-Zamir scheme.

Proposition 9.1 *There exists a sequence of QIM watermarking codes indexed by block length L, such that the value of the mutual-information game (9.2) converges to the unconstrained capacity $C = \frac{1}{2} \ln(1 + D_1/D_2)$. The corresponding asymptotically maxmin solution is $\alpha = \frac{D_1}{D_1 + D_2}$ and $p_{\mathbf{W}} \sim \mathcal{N}(0, D_2 I_L)$.*

Proof: See Appendix E.

10 Conclusion

We have investigated a systematic approach to lattice QIM code design, based on probability-of-error and Bhattacharyya performance measures. For any value of the lattice inflation ("Costa") parameter α, the worst-case attack pdf is obtained as the solution to a convex program; hence globally optimal solutions can be obtained numerically. In some cases, analytical solutions can be found. For instance, if $\alpha \geq \frac{3}{4}$, the worst attack pdf against scalar QIM (with $q = 2$) allocates all of its mass to a small set of values.

We have optimized the parameter α as well. Since the embedder does not know the attack pdf, α is the solution to a minmax problem. We have found that impulsive-noise attacks are very effective at moderate-to-large values of WNR; the suboptimality of conventional AWGN attacks is striking at high WNR's.

Another useful strategy for the embedder is randomized rotation of the QIM lattice. This strategy improves the robustness of the QIM code against surgical attacks. The worst attack pdf is then isotropic.

It appears that making q (the size of the quotient group $\mathcal{C}$) large improves robustness against surgical attacks. This view is supported by the capacity results of Sec. 9, where the input alphabet to the MLAN channel is $\mathcal{V}$ (a continuum) rather than a size-q discrete subset of $\mathcal{V}$. Numerical results by Tzschoppe *et al* [14] also support that view.

Throughout this paper, we have allowed the attacker to use memory as large as the dimension L of the QIM lattice. As discussed below Prop. 8.2, the larger L is, the more advantageous the outcome seems to be for the attacker, even when randomized lattice rotation is allowed. For instance, the error exponent was $O(\frac{\ln L}{L})$ for the STDM example considered. Perhaps this result should not be surprising if we recall that the attacker operates under *average-distortion constraints* and we view the above result in light of the studies in [9, 10]. There, if the attacker's memory is n, he may use a devastating nonergodic strategy: "do nothing" with a fixed probability $1 - \epsilon$, and "kill the signal" with probability ϵ (incurring a large but finite distortion). The resulting probability of error is ϵ, independently of the value of n. The AWGN+Delta attack developed in Sec. 8.2 is somewhat analogous, in that the attacker uses overwhelming power $O(\ln L)$ (vectors $\mathbf{W}$ with total energy $L(\ln L)D_2$) with low probability $O(\frac{1}{\ln L})$. Note that the probability of atypically large distortions (with respect to the expected value) increases dramatically with L. Indeed, using large-deviations analysis one can show that $\frac{1}{n} \log Pr[\frac{1}{n} \sum_{i=1}^{n/L} \|\mathbf{W}_i\|^2 \geq aD_1] = O(1/L)$ for any fixed $a > 1$.

For large L one could argue that by introducing large distortion with excessive probability, the attacker exploits a loophole in the rules of the game. To close this loophole, for large L we may want to replace the average-distortion constraint used throughout this paper with a stronger *maximum-distortion constraint* on each vector $\mathbf{W}$. Mathematically, this equivalent to constraining the support of $p_{\mathbf{W}}$ to a spherical ball $\mathcal{B}$ with radius $\sqrt{LD_2}$, which introduces a new linear equality constraint, $\int_{\mathcal{B}} p_{\mathbf{W}} = 1$, and therefore leads to the exact same kind of optimization algorithms.

Another observation is that our rules allow the attacker to create error patterns that occur in bursts (e.g., the AWGN+Delta attack wipes out entire blocks). This strategy could be circumvented by using interleaving prior to block-QIM embedding; interleaving has the effect of making the attack channel essentially memoryless [32]. The results of Sec. 5 on worst memoryless attacks become all the more relevant in this context.

Acknowledgements. We thank R. Tzschoppe for providing us with a copy of his preprint [14] and calling the paper [12] to our attention. We are particularly grateful to Prof. Pérez-González and Dr. Kalker for carefully reading this manuscript and proposing substantial improvements, and to Ying Wang for improving the implementation of our optimization algorithms.

A Decoding Error Bounds

This appendix presents a brief derivation of coding bounds and motivates the use of the Bhattacharyya parameter (4.16) as a performance metric. The first part of this material is detailed in Gallager's book [17], and the second part extends it to the case of block codes with shared parameters $\{\mathsf{G}(i), 1 \le i \le n_B\}$ at the encoder and decoder.

Consider a message set $\mathcal{M}$ and a set of $|\mathcal{M}|$ codewords associated with the messages in $\mathcal{M}$. Consider a binary hypothesis test between codewords $\mathbf{c}$ and $\mathbf{c}'$ based on the received data $\tilde{\mathbf{y}}^{1:n_B}$. Denote by $n_{cc'}(\mathbf{c}, \mathbf{c}'), 0 \le c, c' < q$, the joint composition of the codewords $\mathbf{c}$ and $\mathbf{c}'$, i.e., the number of positions i where $\mathbf{c}(i) = c$ and $\mathbf{c}'(i) = c'$. The numbers $n_{cc'}(\mathbf{c}, \mathbf{c}')$ sum to n_B. The probability of error for this test is given by

$$\begin{aligned}
P_e(m, m') &= \frac{1}{2} \int_{\mathcal{Y}^{n_B}} \min[p(\tilde{\mathbf{y}}^{1:n_B}|m), p(\tilde{\mathbf{y}}^{1:n_B}|m')]\, d\tilde{\mathbf{y}}^{1:n_B} \\
&= \frac{1}{2} \int_{\mathcal{V}} \cdots \int_{\mathcal{V}} \min \left[\prod_{i=1}^{n_B} p_{\mathbf{c}(i)}(\tilde{\mathbf{y}}(i)), \prod_{i=1}^{n_B} p_{\mathbf{c}'(i)}(\tilde{\mathbf{y}}(i)) \right] d\tilde{\mathbf{y}}(1) \cdots d\tilde{\mathbf{y}}(n_B) \\
&\le \frac{1}{2} \int_{\mathcal{V}} \cdots \int_{\mathcal{V}} \left[\prod_{i=1}^{n_B} p_{\mathbf{c}(i)}(\tilde{\mathbf{y}}(i))\, p_{\mathbf{c}'(i)}(\tilde{\mathbf{y}}(i)) \right]^{1/2} d\tilde{\mathbf{y}}(1) \cdots d\tilde{\mathbf{y}}(n_B) \\
&= \frac{1}{2} \prod_{c,c'=0}^{q-1} \left[\int_{\mathcal{V}} \sqrt{p_c(\tilde{\mathbf{y}})\, p_{c'}(\tilde{\mathbf{y}})}\, d\tilde{\mathbf{y}} \right]^{n_{cc'}(\mathbf{c},\mathbf{c}')} \\
&= \frac{1}{2} \exp \left\{ -L \sum_{c,c'=0}^{q-1} n_{cc'}(\mathbf{c}, \mathbf{c}') B(p_c, p_{c'}) \right\}
\end{aligned}$$

where we have used the inequality $\min(p, q) \le \sqrt{pq}$ which holds for any nonnegative numbers p and q. Using the union bound, we obtain

$$P_e \le \frac{|\mathcal{M}| - 1}{2} \max_{\mathbf{c},\mathbf{c}'} \exp \left\{ -L \sum_{c,c'=0}^{q-1} n_{cc'}(\mathbf{c}, \mathbf{c}') B(p_c, p_{c'}) \right\}. \tag{A.1}$$

Consider codes such that

$$n_{cc'}(\mathbf{c}, \mathbf{c}') \sim \begin{cases} \frac{1-\overline{d}_{\min}}{q} n_B & : \ c = c' \\ \frac{\overline{d}_{\min}}{q(q-1)} n_B & : \ c \ne c' \end{cases}$$

i.e, they have uniform distribution over the alphabet $\{0, \cdots, q-1\}$, normalized minimum distance $\overline{d}_{\min}$, and codewords pairs have constant joint composition. Then (A.1) simplifies into

$$P_e \le \frac{|\mathcal{M}| - 1}{2} \exp \left\{ -n \overline{d}_{\min} \frac{1}{q(q-1)} \sum_{c \ne c'} B(p_c, p_{c'}) \right\}.$$

Assume now that a parameter $\mathsf{G}(i) \in \mathcal{G}$ is generated for each block i, by drawing independently from a distribution μ on $\mathcal{G}$. The parameters are shared with the decoder; denote by $p_c(\tilde{y}|\mathsf{G})$ the conditional probability distribution at the output of the channel. The calculation above extends in a straightforward manner to this scenario. Let $B(p_c, p_{c'}|\mathsf{G})$ denote the Bhattacharyya distance between $p_c(\tilde{y}|\mathsf{G})$ and $p_{c'}(\tilde{y}|\mathsf{G})$. Then

$$\begin{aligned} P_e(m,m') &= \frac{1}{2}\int_{\mathcal{V}}\int_{\mathcal{G}}\cdots\int_{\mathcal{V}}\int_{\mathcal{G}} \min\left[\text{`}\prod_{i=1}^{n_B} p_{\mathbf{c}(i)}(\tilde{\mathbf{y}}(i)|\mathsf{G}(i)), \prod_{i=1}^{n_B} p_{\mathbf{c}'(i)}(\tilde{\mathbf{y}}(i)|\mathsf{G}(i))\right] \\ &\qquad \times d\tilde{\mathbf{y}}(1)\, d\mu(\mathsf{G}(i)) \cdots d\tilde{\mathbf{y}}(n_B)\, d\mu(\mathsf{G}(n_B)) \\ &\leq \frac{1}{2}\exp\left\{-L\sum_{c,c'=0}^{q-1} n_{cc'}(\mathbf{c},\mathbf{c}')\int_{\mathcal{G}} B(p_c,p_{c'}|\mathsf{G})\, d\mu(\mathsf{G})\right\} \end{aligned}$$

and so all subsequent derivations hold with $\int_{\mathcal{G}} B(p_c, p_{c'}|\mathsf{G})\, d\mu(\mathsf{G})$ in place of $B(p_c, p_{c'})$.

B Proof of Lemma 4.1

(i) It suffices to prove that the function $P_e(\alpha, p_{\mathbf{W}})$ is concave in $p_{\mathbf{W}}$ over the set of n-dimensional pdf's. For any n-dimensional pdf's $p_{\mathbf{W}}$, $p'_{\mathbf{W}}$, and constant $\theta \in [0,1]$, we prove that

$$P_e(\alpha, \theta p_{\mathbf{W}} + (1-\theta)p'_{\mathbf{W}}) \geq \theta P_e(\alpha, p_{\mathbf{W}}) + (1-\theta)P_e(\alpha, p'_{\mathbf{W}}). \tag{B.1}$$

Define a binary random variable $T \in \{0,1\}$ independent of all other random variables in the problem, and such that $Pr[T=0] = \theta$. Given any two attack channels $p_{\mathbf{W}}$ and $p'_{\mathbf{W}}$, suppose the attacker observes T and selects $p_{\mathbf{W}}$ if $T=0$ and $p'_{\mathbf{W}}$ otherwise. Equivalently the attacker selects the noise vector $\mathbf{w}$ according to the mixture pdf $\theta p_{\mathbf{W}} + (1-\theta)p'_{\mathbf{W}}$. If T is unknown to the decoder, the probability of error is $P_e(\alpha, \theta p_{\mathbf{W}} + (1-\theta)p'_{\mathbf{W}})$. However, if T is known to the decoder, the probability of error is $\theta P_e(\alpha, p_{\mathbf{W}}) + (1-\theta)P_e(\alpha, p'_{\mathbf{W}})$. Since the decoder has more information in the latter case, probability of error cannot exceed that in the first case. This proves (B.1).

(ii) Since $q=2$, the functional $\exp\{-Lb(\alpha, p_{\mathbf{W}})\} = \int_{\mathcal{V}} p_0 \psi(p_1/p_0)$ is a negative f-divergence. The claim is then a direct consequence of Lemma 4.1 in [28, p. 448]: given two pdf's p and q defined over a common domain, the negative f-divergence $\int p\, \psi(\frac{q}{p})$ is concave in the pair (p,q). In our problem, (p,q) are subject to linear constraints, and the common domain is $\mathcal{V}$.

(iii) When $q=2$, the Bhattacharyya distance $b(\alpha, p_{\mathbf{W}})$ is a convex decreasing function of the negative f-divergence of Part (ii). Since that negative f-divergence is concave in $p_{\mathbf{W}}$, $b(\alpha,\cdot)$ is convex [24]. When $q>2$, $b(\alpha,\cdot)$ is a sum of convex functions and is therefore convex.

(iv) The function $-\frac{1}{n}\ln P_e(\alpha, p_{\mathbf{W}})$ is a convex decreasing function of $P_e(\alpha, p_{\mathbf{W}})$, which is itself concave in $p_{\mathbf{W}}$ (as proved in Part (i)). The functional $-\frac{1}{n}\ln P_e(\alpha,\cdot)$ is therefore convex. □

C Proof of Proposition 5.3

Because $\alpha = 1$ and the support set of p_W is limited to $[-\frac{\Delta}{2}, \frac{\Delta}{2}]$, we have $\overset{\circ}{p}_V = \overset{\circ}{p}_W$. The attacker's minimization problem (5.6) may be written as

$$\text{Minimize } \beta(1, p_W) = -\ln \int_{-\Delta/2}^{\Delta/2} \overset{\circ}{p}_W(w - \Delta/4)\, \psi\left(\frac{\overset{\circ}{p}_W(w+\Delta/4)}{\overset{\circ}{p}_W(w-\Delta/4)}\right) dw \tag{C.1}$$

subject to

$$\int_{-\Delta/2}^{\Delta/2} w^2 p_W(w)\,dw \;\leq\; D_2, \tag{C.2}$$

$$\int_{-\Delta/2}^{\Delta/2} p_W(w)\,dw \;=\; 1. \tag{C.3}$$

For WNR $\leq \frac{4}{3}$, the optimization problem admits a straightforward solution:

$$p_W(w) = \frac{1}{2}[\delta(w-\Delta/4) + \delta(w+\Delta/4)]$$

(the "two-delta" solution), which yields $\beta(1, p_W) = -\ln\psi(1) = \beta_{\min}$. The distortion constraint is inactive for all WNR $< \frac{4}{3}$.

When WNR $> \frac{4}{3}$, the distortion constraint (C.2) is active, and the proof proceeds as follows. Define the ratio

$$r(w) = \frac{p_W(\Delta/2 - w)}{p_W(w)}, \quad 0 \leq w \leq \frac{\Delta}{4}. \tag{C.4}$$

Due to Prop. 5.1, the optimization may be restricted to symmetric noise pdf's. The problem (C.1)–(C.3) may then be written in the equivalent form

$$\text{Minimize } \beta(1, p_W) = \tilde{\beta}(p_W, r) \triangleq -\ln\left\{4\int_0^{\Delta/4} p_W(w)\,\psi(r(w))\,dw\right\} \tag{C.5}$$

where p_W and r are subject to the constraints

$$2\int_0^{\Delta/4} [w^2 + (\Delta/2 - w)^2 r(w)]\,p_W(w)\,dw \;=\; D_2, \tag{C.6}$$

$$2\int_0^{\Delta/4} [1 + r(w)]\,p_W(w)\,dw \;=\; 1. \tag{C.7}$$

Denote by β^* the minimum of (C.5). The minimum over all pdf's is equal to the infimum over all discrete distributions. Denoting by $\{w_i, i \in \mathcal{I}\}$ the restriction of the support set of discrete p_W to the open interval $(0, \Delta/4)$, we write

$$\beta^* = \inf_{\mathcal{I}} \inf_{\mathbf{p},\mathbf{r},\mathbf{w}} -\ln\left\{4\sum_{i\in\mathcal{I}} p_i\,\psi(r_i)\right\} \tag{C.8}$$

where the infima are subject to

$$2\sum_{i\in\mathcal{I}} [w_i^2 + (\Delta/2 - w_i)^2 r_i]\,p_i \;=\; D_2, \tag{C.9}$$

$$2\sum_{i\in\mathcal{I}} [1 + r_i]\,p_i \;=\; 1. \tag{C.10}$$

The first infimum in (C.8) is over all discrete subsets of the open interval $(0, \Delta/4)$, and the sequence $\mathbf{w}$ has components $w_i \in (0, \Delta/4)$. The sequences $\mathbf{p}$ and $\mathbf{r}$ have components $p_i = p_W(w_i)$ and $r_i = r(w_i)$, respectively. Observe now that the optimization problem is separable in the components

of $\mathcal{I}$ and invariant to permutations of the joint sequence $(\mathbf{p}, \mathbf{r}, \mathbf{w})$. Therefore, owing to the convexity of the optimization problem, there is no loss in optimality in requiring that all p_i, r_i and w_i be independent of i. Let p, r and w denote the corresponding values of $|\mathcal{I}|p_i$, r_i, and w_i, respectively. We obtain

$$\beta^* = \inf_{p,r,w\geq 0} -\ln[4p\,\psi(r)] \tag{C.11}$$

subject to

$$2[w^2 + (\Delta/2 - w)^2 r]\,p = D_2, \tag{C.12}$$

$$2[1+r]\,p = 1. \tag{C.13}$$

In other words, without loss of optimality, we may restrict our attention to index sets $\mathcal{I}$ made of a single element. The resulting p_W will be a distribution with support at four points $\pm w$ and $\pm(\frac{\Delta}{2} - w)$.

Next, observe that the optimization over w may be absorbed in the distortion constraint:

$$\beta^* = \inf_{p,r\geq 0} -\ln[4p\,\psi(r)] \tag{C.14}$$

subject to

$$2 \inf_{0<w<\Delta/4} [w^2 + (\Delta/2 - w)^2 r]\,p = D_2, \tag{C.15}$$

$$2[1+r]\,p = 1. \tag{C.16}$$

The optimization over w in (C.15) is a simple quadratic problem whose solution is

$$w = \frac{\Delta}{2}\,\frac{r}{r+1} \tag{C.17}$$

$$\Rightarrow \quad w^2 + (\Delta/2 - w)^2 r = \left(\frac{\Delta}{2}\right)^2 \frac{r}{r+1} = 3\,D_1\,\frac{r}{r+1}.$$

Hence

$$\beta^* = \inf_{p,r\geq 0} -\ln[4p\,\psi(r)] \tag{C.18}$$

subject to

$$\frac{r}{r+1}\,p = \frac{1}{6\,\mathrm{WNR}}, \tag{C.19}$$

$$2[1+r]\,p = 1. \tag{C.20}$$

The system (C.19), (C.20) has a unique solution. Putting $\gamma = \sqrt{1 - \frac{4}{3\,\mathrm{WNR}}}$, we obtain

$$p = \frac{1+\gamma}{4} \quad \text{and} \quad r = \frac{1-\gamma}{1+\gamma} \leq 1. \tag{C.21}$$

Substituting these values in (C.17), we obtain $w = \frac{\Delta}{4}(1-\gamma)$. The solution does not depend on the function $\psi(\cdot)$; however the value of the game does.

In the Bhattacharyya case, we have $\psi(r) = \sqrt{r}$, hence (C.21) yields

$$\beta^* = -\ln[4p\sqrt{r}] = -\ln\sqrt{1-\gamma^2} = -\ln\sqrt{\frac{4}{3\,\mathrm{WNR}}}.$$

In the P_e case, we have $\psi(r) = \frac{r}{2}$ because $r \leq 1$, and thus (C.21) yields

$$P_e = e^{-\beta^*} = 2pr = \frac{1-\gamma}{2}.$$

This concludes the proof. In (5.7), we have used the notation w^* for w and $\frac{1-a}{2}$ for p. □

D Proof of Proposition 5.4

Under the 3-delta attack, for all $\alpha \geq \frac{1}{2}$, p_V is made of nonoverlapping rectangular pulses at locations 0 and $\Delta/2$:

$$p_V(v) = \frac{1}{(1-\alpha)\Delta}\left[(1-a)\,1_{\{|v|\leq(1-\alpha)\Delta/2\}} + a\,1_{\{|\,|v|-\Delta/2|\leq(1-\alpha)\Delta/2\}}\right].$$

Owing to (5.4), we have

$$\sqrt{p_0(\tilde{y})p_1(\tilde{y})} = \frac{\sqrt{a(1-a)}}{(1-\alpha)\Delta}\,1_{\{|\,|\tilde{y}|-\Delta/4|\leq(1-\alpha)\Delta/2\}}.$$

The Bhattacharyya distance $b(\alpha, p_W^{(\alpha)}) = -\ln\int\sqrt{p_0 p_1} = -\ln(2\sqrt{a(1-a)})$ is decreasing in a and therefore also in α. Its maximum in the range $[\frac{1}{2}, 1]$ is achieved at $\alpha = \frac{1}{2}$; in this case, $a = \frac{1}{12\,\mathrm{WNR}}$.

For all $\alpha < \frac{1}{2}$, we have $a < \frac{1}{12\,\mathrm{WNR}}$, and the rectangular pulses that make up p_V overlap around $\pm\Delta/4$. The support set of p_V is the whole range $[-\Delta/2, \Delta/2]$:

$$p_V(v) = \frac{1}{(1-\alpha)\Delta}\left[(1-a)\,1_{\{|v|\leq\Delta\alpha/2\}} + 1_{\{|\,|v|-\Delta/4|\leq\Delta(1-2\alpha)/4\}} + a\,1_{\{|\,|v|-\Delta/2|\leq\Delta\alpha/2\}}\right].$$

Elementary calculations yield $b(\alpha, p_W^{(\alpha)}) = -\ln(1-2\alpha+(2\alpha)2\sqrt{a(1-a)})$. and

$$\frac{d}{d\alpha}b(\alpha, p_W^{(\alpha)}) = -e^{b(\alpha,p_W^{(\alpha)})}\left[-2 + \frac{8\alpha}{\sqrt{3\,\mathrm{WNR}}}\,\gamma\left(\frac{\alpha^2}{3\mathrm{WNR}}\right)\right]$$

where $\gamma(x) = (1-x)^{1/2}(1-\frac{x}{2(1-x)}) \leq 1$. For $\mathrm{WNR} \geq \frac{4}{3}$ and $\alpha \leq \frac{1}{2}$, the above derivative is positive, and thus $b(\alpha, p_W^{(\alpha)})$ is increasing in α. Its maximum in the range $[0, \frac{1}{2}]$ is achieved at $\alpha = \frac{1}{2}$.

To summarize, the minimum of the function $b(\alpha, p_W^{(\alpha)})$ over $0 \leq \alpha \leq 1$ is achieved by $\alpha = \frac{1}{2}$ for all $\mathrm{WNR} \geq \frac{4}{3}$. The claim follows by applying (5.9) and observing that the right side of (5.10) is equal to $b(\frac{1}{2}, p_W^{(1/2)})$. □

E Proof of Proposition 9.1

To explicitly indicate the dependency of the coarse lattice Λ on L, we use the notation $\Lambda = \Lambda^{(L)}$. The normalized second moment of $\Lambda^{(L)}$ is defined as [20, 6]

$$G(\Lambda^{(L)}) = |\mathcal{V}|^{-2/L}D_1 \geq \frac{1}{2\pi e}.$$

There exists a sequence of lattices $\Lambda^{(L)}$ that are *good for quantization*, in the sense that $G(\Lambda^{(L)}) \downarrow \frac{1}{2\pi e}$ as $L \to \infty$ [21] [8]. Given such a sequence of lattices, define

$$C_L(\alpha, p_{\mathbf{W}}) \triangleq \frac{1}{L}I(\mathbf{Z}; \tilde{\mathbf{Y}}) = \frac{1}{L}\ln|\mathcal{V}| - \frac{1}{L}h(\mathbf{V}) = \frac{1}{2}\ln\frac{D_1}{G(\Lambda^{(L)})} - \frac{1}{L}h(\mathbf{V}).$$

[8] Loosely speaking, their Voronoi cells are "asymptotically spherical".

We need to prove that

$$\lim_{L\to\infty} \max_{\alpha} \min_{p_{\mathbf{W}}} C_L(\alpha, p_{\mathbf{W}}) = \frac{1}{2}\ln\left(1+\frac{D_1}{D_2}\right) \tag{E.1}$$

in which the asymptotically maxmin solution is $\alpha = \frac{D_1}{D_1+D_2}$ and $p_{\mathbf{W}} \sim \mathcal{N}(0, D_2 I_L)$. Owing to our choice of $\Lambda^{(L)}$, the claim (E.1) is equivalent to

$$\lim_{L\to\infty} \min_{\alpha} \max_{p_{\mathbf{W}}} \frac{1}{L} h(\mathbf{V}) = \frac{1}{2}\ln\left(2\pi e \frac{D_1 D_2}{D_1+D_2}\right). \tag{E.2}$$

The average variance of the components of $\mathbf{V} = \tilde{\mathbf{E}} + \tilde{\mathbf{W}} \bmod \Lambda$ in (4.3) is given by

$$\sigma_v^2 \triangleq \frac{1}{L}\mathbb{E}\|\mathbf{V}\|^2 \le \frac{1}{L}\mathbb{E}\|\tilde{\mathbf{E}}+\tilde{\mathbf{W}}\|^2 = (1-\alpha)^2 D_1 + \alpha^2 D_2 \le \frac{D_1 D_2}{D_1+D_2}$$

where the latter upper bound is achieved by $\alpha = \frac{D_1}{D_1+D_2}$. Let $\mathbf{V}_G$ be a Gaussian random vector with i.i.d. components and average variance σ_v^2. We have [31]

$$\frac{1}{L}h(\mathbf{V}) \le \frac{1}{L}h(\mathbf{V}_G) = \frac{1}{2}\ln(2\pi e\sigma_v^2) \le \frac{1}{2}\ln\left(2\pi e\frac{D_1 D_2}{D_1+D_2}\right) \quad \forall L, \alpha, p_{\mathbf{W}}. \tag{E.3}$$

The Gaussian upper bound (E.3) on $\frac{1}{L}h(\mathbf{V})$ is not achievable for any finite L, because $\tilde{\mathbf{E}}$ is uniformly distributed over $(1-\alpha)\mathcal{V}$ and therefore $\mathbf{V}$ cannot be Gaussian. However Erez and Zamir [6] proved that this upper bound is asymptotically achievable when $p_{\mathbf{W}} \sim \mathcal{N}(0, D_2 I_L)$. Choosing the MMSE value $\alpha = \frac{D_1}{D_1+D_2}$, they obtained

$$\lim_{L\to\infty} \min_{\alpha} \frac{1}{L} h(\mathbf{V}) = \frac{1}{2}\ln\left(2\pi e \frac{D_1 D_2}{D_1+D_2}\right). \tag{E.4}$$

From (E.3) and (E.4), we conclude that $p_{\mathbf{W}} \sim \mathcal{N}(0, D_2 I_L)$ achieves (E.2), i.e., becomes the worst noise when $L \to \infty$ [9]. □

References

[1] B. Chen and G. W. Wornell, "An information–theoretic approach to the design of robust digital watermarking systems," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, Phoenix, AZ, March 1999, pp. 2061—2064.

[2] B. Chen and G. W. Wornell, "Quantization index modulation methods: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Information Theory*, vol. 47, no. 4, pp. 1423—1443, May 2001.

[3] M. Kesal, M. K. Mıhçak, R. Kötter, and P. Moulin, "Iteratively decodable codes for watermarking applications," in *Proc. 2nd Symposium on Turbo Codes and Related Topics*, Brest, France, Sep. 2000.

[4] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Information Theory*, vol. 48, pp. 1250—1276, June 2002.

[5] J. J. Eggers, R. Bäuml, R. Tzschoppe, and B. Girod, "Scalar Costa scheme for information embedding," *IEEE Trans. Sig. Proc.*, vol. 51, no. 4, pp. 1003—1019, April 2003.

[9] It may be verified that a uniform distribution over the scaled Voronoi cell $\sqrt{D_2/D_1}\mathcal{V}$ would have the same asymptotic performance.

[6] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+SNR)$ on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Information Theory*, vol. 50, no. 10, pp. 2293—2314, Oct. 2004.

[7] P. Moulin and R. Koetter, "Data-Hiding Codes," *Proceedings IEEE*, Vol. 93, No. 12, pp. 2081—2127, Dec. 2005.

[8] T. Liu, P. Moulin and R. Koetter, "On Error Exponents of Modulo Lattice Additive Noise Channels," *IEEE Trans. on Information Theory*, Vol. 52, No. 2, pp. 454–471, Feb. 2006.

[9] A. S. Cohen and A. Lapidoth, "The Gaussian Watermarking Game," *IEEE Trans. Information Theory*, Vol. 48, No. 6, pp. 1639—1667, June 2002.

[10] P. Moulin and J. A. O'Sullivan, "Information–theoretic analysis of information hiding," *IEEE Trans. Information Theory*, vol. 49, no. 3, pp. 563—593, March 2003.

[11] A. K. Goteti, "Optimal Strategies for Private and Public Watermarking Games," M.S. Thesis, ECE Department, University of Illinois at Urbana-Champaign, Dec. 2004.

[12] F. Pérez-González, "The Importance of Aliasing in Structured Quantization Index Modulation," *Proc. Int. Workshop on Digital Watermarking*, Seoul, Korea, Nov. 2003.

[13] J.E. Vila-Forcén, S. Voloshynovskiy, O. Koval, F. Pérez-González and T. Pun, "Worst Case Additive Attack Against Quantization-Based Watermarking Techniques", *Proc. IEEE Int. Workshop on Multimedia Signal Processing*, Siena, Italy, Sep. 2004.

[14] R. Tzschoppe, R. Bäuml, R. Fischer, A. Kaup and J. Huber, "Additive non-Gaussian Attacks on the Scalar Costa Scheme (SCS)," *Proc. SPIE*, San Jose, CA, Jan. 2005.

[15] S. Shamai and S. Verdú, "Worst-Case Power-Constrained Noise for Binary-Input Channels," *IEEE Trans. on Information Theory*, Vol. 38, No. 5, pp. 1494—1511, 1992.

[16] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels. II", *Information and Control*, Vol. 10, pp. 522—552, 1967.

[17] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, 1968.

[18] R. E. Blahut, "Composition Bounds for Channel Block Codes," *IEEE Trans. Information Theory*, Vol. 23, No. 6, pp. 656–674, Nov. 1977.

[19] F. Jelinek, "Evaluation of Expurgated Bound Exponents," *IEEE Trans. Information Theory*, vol. 14, No. 3, pp. 501–505, May 1968.

[20] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, 3rd ed. New York: Springer-Verlag, 1999.

[21] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Information Theory*, vol. 42, pp. 1152—1159, July 1996.

[22] H. V. Poor, *An Introduction to Detection and Estimation Theory.* New York: Springer-Verlag, 1994.

[23] A. Bhattacharyya, "On a Measure of Divergence Between Two Statistical Populations Defined by Their Probability Distributions," *Bulletin of the Calcutta Mathematical Society*, vol. 35, No. 3, pp. 99—109, Sep. 1943.

[24] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, UK, 2004.

[25] D. L. Luenberger, *Optimization by Vector Space Methods.* New York: Wiley, 1969.

[26] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. Royal Statistical Society*, series B, vol. 28, pp. 132—142, 1966.

[27] I. Csiszár, "Information-Type Distance Measures and Indirect Observations," *Stud. Sci. Math. Hungar.*, Vol. 2, pp. 299–318, 1967.

[28] I. Csiszár and P. C. Shields, Information Theory And Statistics: A Tutorial, *Foundations and Trends in Communications and Information Theory*, Vol. 1, No. 4, pp. 417—528, 2004. Available from `htpp://www.renyi.hu/~csiszar`.

[29] P. Moulin, A. K. Goteti, and R. Koetter, "Optimal sparse-QIM codes for zero-rate blind watermarking," *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, Montreal, Canada, May 2004.

[30] The TOMLAB optimization environment, 2004, `http://www.tomlab.biz`.

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

[32] A. J. Viterbi, "Spread Spectrum Communications – Myths and Realities," *IEEE Communications Magazine*, Vol. 17, No. 3, pp. 11—18, May 1979.

	Scheme	L	α_{opt}	Attack ($p_{\mathbf{W}}$)	$b(\alpha, p_{\mathbf{W}})$
Sec. 5	scalar QIM	1	≈ 0.5	AWGN	0.232
			≈ 0.5	worst p_W	0.214
Sec. 6	cubic QIM	2	≈ 0.5	worst $p_{\mathbf{W}}$	0.195
Sec. 7			≈ 0.5	worst isotropic	0.235
Sec. 8	scalar STDM	10	≈ 1	AWGN	0.375
			≈ 1	AWGN+Delta	0.255

Table 1: Detection performance for $q = 2$, WNR $= 1$.

	Scheme	L	α_{opt}	Attack ($p_{\mathbf{W}}$)	$b(\alpha, p_{\mathbf{W}})$
Sec. 5	scalar QIM	1	≈ 0.5	AWGN	0.177
			≈ 0.5	worst p_W	0.154
Sec. 6	hex QIM	2	≈ 0.5	AWGN	0.218
			≈ 0.5	worst $p_{\mathbf{W}}$	0.201

Table 2: Detection performance for $q = 3$, WNR $= 1$.

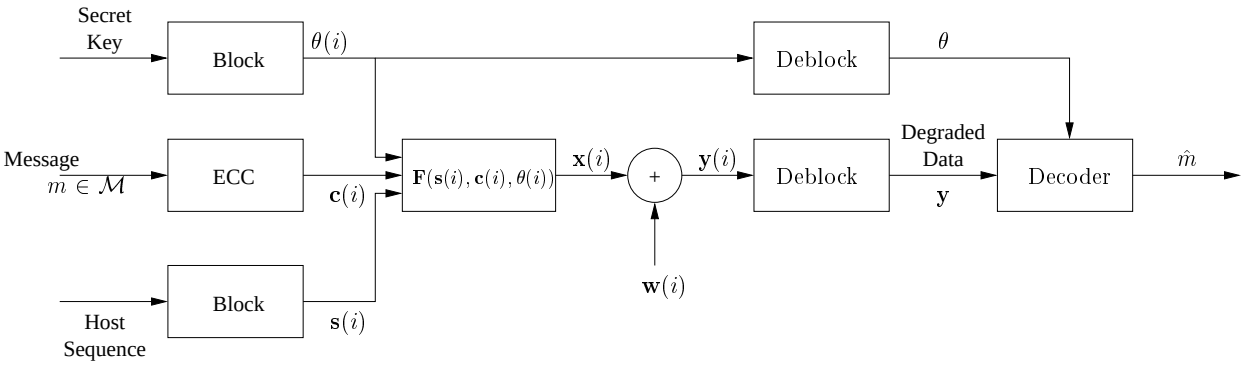
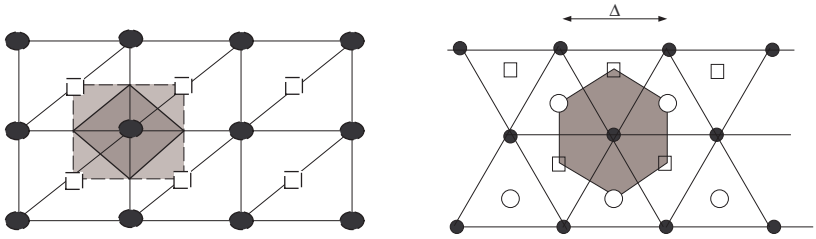


Figure 1: Communication model for watermarking. The encoder is a two-stage encoder.



(a) quincunx lattice, $q = 2$ (b) hexagonal lattice, $q = 3$

Figure 2: Nested two-dimensional lattices. The coarse lattice Λ is the set of heavy dots and its cosets are represented by squares and circles. The lightly shaded region is $\mathcal{V}$, the Voronoi cell of Λ.

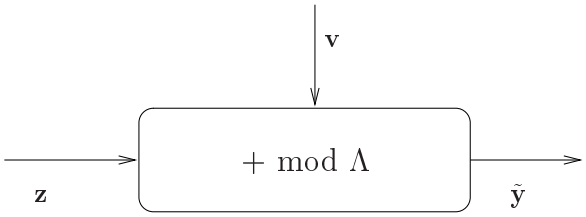


Figure 3: Modulo Lattice Additive Noise channel.

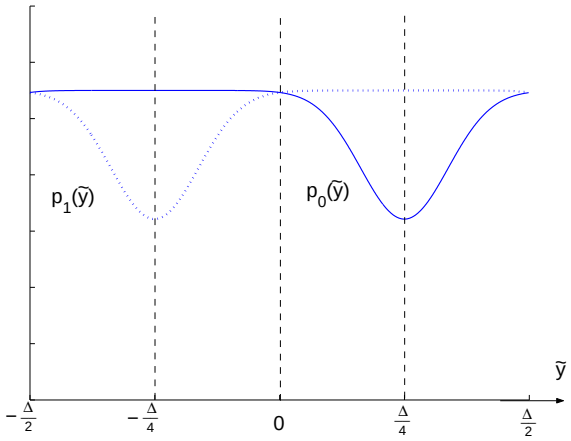


Figure 4: Example of $p_0(\tilde{y})$ and $p_1(\tilde{y})$. Here WNR $= 0.1$, $\alpha = 0.09$, and W is Gaussian.

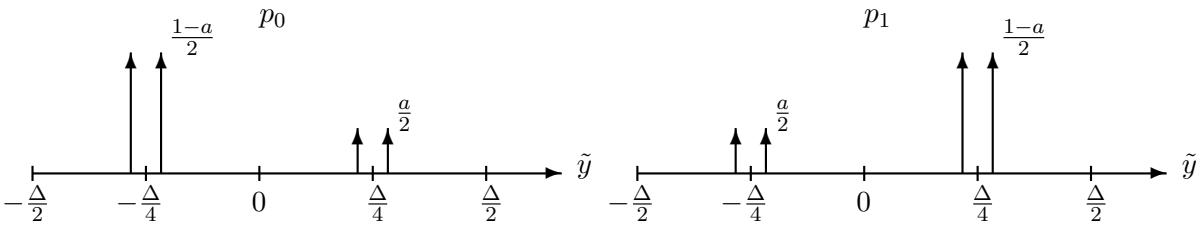


Figure 5: Rival pdf's $p_0(\tilde{y})$ and $p_1(\tilde{y})$ when $\alpha = 1$ and the worst p_W is used.

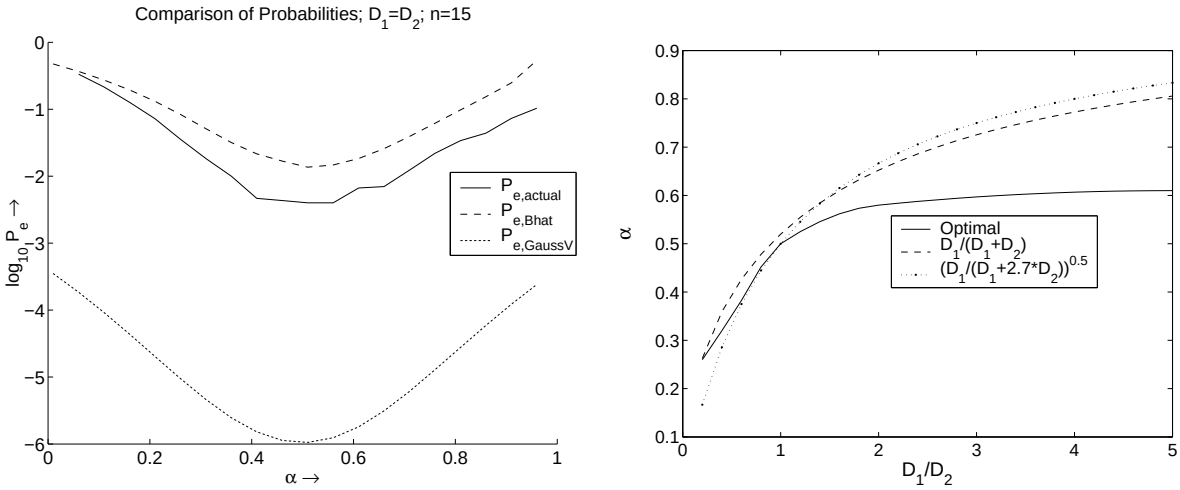
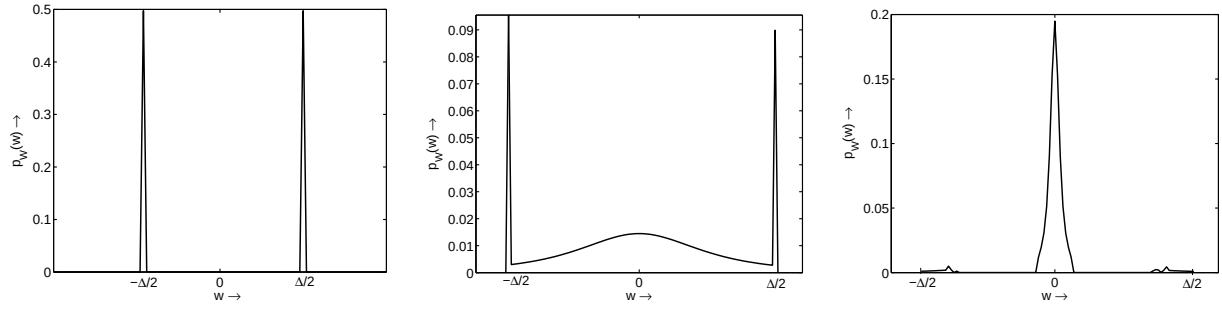


Figure 6: (a) Actual P_e and Bhattacharyya bound on P_e for WNR $= 1, n = 15$, and AWGN attack on scalar QIM ($L = 1$). (b) Optimal α *vs* WNR for worst-case p_W.

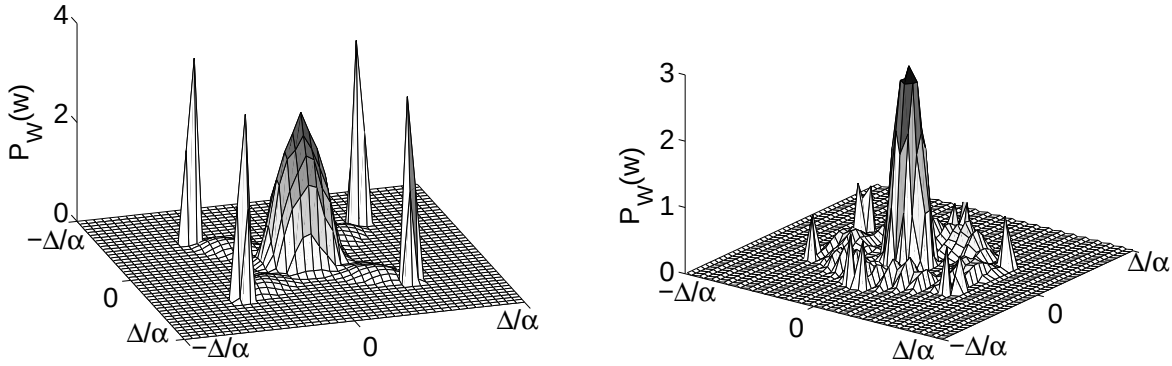


(a) WNR $= 0.2$, $\alpha_{opt} \approx 0.25$, $b^* = 0.030$

(b) WNR $= 1$, $\alpha_{opt} \approx 0.5$, $b^* = 0.214$

(c) WNR $= 5$, $\alpha_{opt} \approx 0.6$. $b^* = 0.980$

Figure 7: Worst-case memoryless noise pdf p_W against scalar QIM. $b^* = b(\alpha_{opt}, p_W)$ denotes the value of the Bhattacharyya game.



(a) (b)

Figure 8: Worst noise pdf $p_{\mathbf{W}}$ for (a) cubic QIM scheme of Fig. 2(a) with $q = 2$, and (b) hexagonal QIM scheme of Fig. 2(c) with $q = 3$. In both cases, WNR $= 1$.

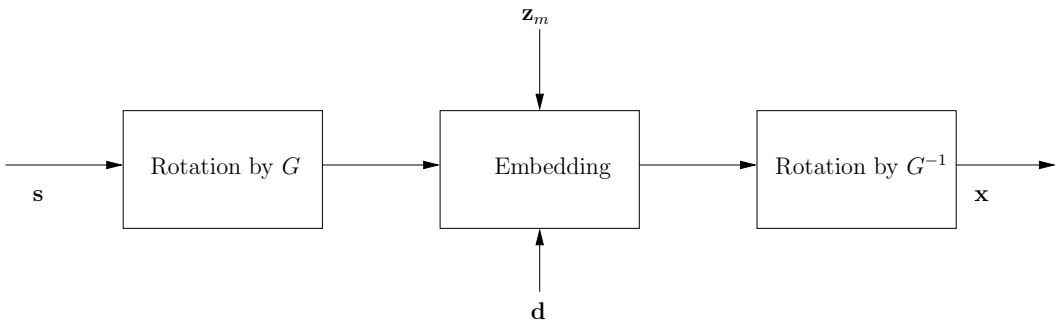
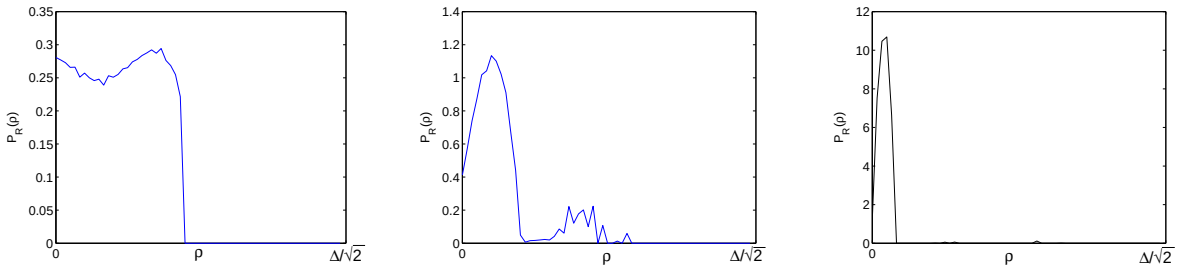


Figure 9: QIM embedding of message $m \in \{0, 1, \cdots, q-1\}$ in length-L block $\mathbf{s}$ using randomized lattice rotation and translation.



(a) WNR $= 0.2$, optimal $\alpha = 0.24$, $b = 0.032$

(b) WNR $= 1$, optimal $\alpha = 0.48$, $b = 0.235$

(c) WNR $= 5$, optimal $\alpha = 0.53$, $b = 0.713$

Figure 10: Worst-case radial noise pdf $p_R(\rho)$ for quincunx lattice code of Fig. 2(a) ($L = 2$, $q = 2$).