Two-Level Quantized Soft-Output Demodulation of QAM Signals With Gray Labeling: A Geometric Approach

Gianluigi Ferrari, *Senior Member, IEEE*, and Umberto Amadei

***Abstract*— We propose a geometric approach to the design of a look-up table-based two-level quantized soft-output (SO) demodulator for coded schemes with quadrature amplitude modulation (QAM) signals with Gray bit labeling. This allows to derive quantized bit logarithmic likelihood ratios (LLRs) directly from the observables, without requiring the actual computation, followed by quantization, of LLRs at the output of the demodulator. The proposed approach is applied to a bit-interleaved coded modulation (BICM) scheme. The obtained results show a limited performance loss with respect to unquantized SO demodulation.**

***Index Terms*— Demodulation, quantization, quadrature amplitude modulation (QAM), bit-interleaved coded modulation (BICM).**

I. Introduction

CURRENT digital communication systems (wireless and wired) are looking with ever increasing interest to the use of high-order quadrature amplitude modulation (QAM) as a straightforward way to increase the spectral efficiency [1], [2]. In this context, separating demodulation and decoding, as in bit-interleaved coded modulation (BICM) schemes [3], [4], is very attractive, as it allows to keep the receiver complexity limited. At the receiver side, in fact, soft-output (SO) demodulation allows the use of advanced decoding techniques (e.g., graph-based), making thus possible the use of powerful channel codes at the transmitter side. In order to further limit the demodulation/decoding complexity, the use of finite-precision signal representation is very attractive [5]. Quantized SO demodulation for BICM schemes in the presence of fading channels is investigated in [6]. In [7], the authors consider Reed-Solomon coded QAM systems and propose an efficient decoding algorithm where the reliabilities of the codewords are derived, through geometric considerations, from the relative positions (in the QAM constellation) of the observables.

In this letter, we propose a novel geometric approach to *binary* quantized SO demodulation of QAM signals with Gray bit labeling, which allows to obtain two-level quantized reliability values directly from the observables, rather than computing LLRs and, then, quantizing them. In particular, the values of the quantized soft output values depend on a single parameter, which is properly optimized. The proposed approach is then applied to an illustrative BICM scheme. The obtained results show that our approach, amenable to an efficient look-up table-based implementation of the demodulator, entails a limited performance loss with respect to the equivalent BICM scheme with unquantized SO demodulation.

Manuscript received June 12, 2016; accepted July 16, 2016. Date of publication July 19, 2016; date of current version October 7, 2016. The associate editor coordinating the review of this letter and approving it for publication was N. Zlatanov.

G. Ferrari is with the Wireless Ad-hoc and Sensor Networks Laboratory, Department of Information Engineering, University of Parma, Parma 43124, Italy (e-mail: gianluigi.ferrari@unipr.it).

U. Amadei is with Bertel Srl - Tesmec Automation Group, Parma 43036, Italy (e-mail: umberto.amadei@bertel.it).

Digital Object Identifier 10.1109/LCOMM.2016.2592963

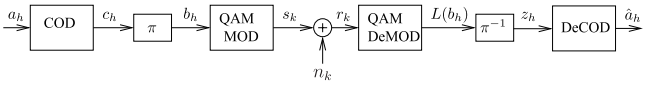

Fig. 1. System model.

II. Soft-Output (Bit-Level) QAM Demodulation and BICM

The considered (discrete time) system model is shown in Fig. 1. At the transmitter side, a sequence of information bits $\{a_h\}$ is coded through a binary channel code, as in classical BICM schemes [3], [4]. After bit interleaving, the coded bits $\{b_h\}$ are mapped[1] into QAM symbols $\{s_k\}$. Considering the equivalent discrete time channel model, the (complex) observable at the receiver side can be expressed as follows:

$$r_k = s_k + n_k$$

where: $k = 1, 2, \ldots$ is the time epoch; $s_k \in \mathcal{S}$ is the QAM symbol, where $\mathcal{S}$ is the set of complex symbols in the considered QAM constellation; and $\{n_k\}$ is a complex additive Gaussian noise process, with independent real and imaginary parts, each with distribution $\mathcal{N}(0, \sigma)$. We consider a square QAM constellation and denote the horizontal/vertical distance between any pair of adjacent QAM symbols as d, i.e., we assume that QAM symbols have generic coordinates $\{\pm i, \pm j\}$, where $i, j \in \{d/2, 3d/2, \ldots, (2\sqrt{M/4} - 1)d/2\}$. The proposed approach can be applied also in the presence of frequency non-selective fading, i.e., if the observable can be expressed as $r_k = f_k s_k + n_k$, where $\{f_k\}$ are independent and identically distributed complex random variables, each with distribution $\mathcal{N}_{\mathbb{C}}(0, \sigma_f)$. Under the assumption of perfect fading knowledge, the quantized demodulation strategy presented in the following can be directly applied by considering the modified observables $\{r'_k = r_k / f_k\}$. Extensions of our approach to scenarios with frequency-selective fading or in the presence of joint demodulation/fading estimation are interesting research directions which go beyond the scope of this paper.

In the presence of M-QAM, each symbol is associated with $m = \log_2 M$ bits. In particular, we denote as $(b_1, b_2, \ldots, b_m)$ the m bits associated with a generic M-QAM symbol s. The specific mapping of sequences of m bits into the constellation points identifies the QAM labeling. In this work, we will focus on Gray labeling, i.e., a labeling according to which adjacent constellation symbols (at minimum distance) are associated with bit sequences which differ by only one bit [8]. This

[1] Note that the time index h of the sequence $\{b_h\}$ is different from the index k of the QAM symbols $\{s_k\}$: the exact relation depends on the specific QAM symbol mapping.

labeling exists for square QAM constellations and is a key ingredient for the proposed quantization strategy.

In the presence of zero-mean circularly symmetric Gaussian noise, considering the Voronoi tesellation associated with the QAM constellation of interest [9], *hard-output* (HO) demodulation can be interpreted as the selection of the QAM symbol associated with the Voronoi region within which the received observable r falls—the time epoch of the observable is omitted for notational simplicity. In particular, the estimated symbol at the output of a HO demodulator can be expressed as follows:

$$s_{\mathrm{HO}} = \underset{s\in\mathcal{S}}{\operatorname{argmin}} |r-s|^2. \tag{1}$$

In the presence of bit-level SO demodulation, logarithmic likelihood ratios (LLRs) are generated for all bits associated with a QAM symbol. In particular, the LLR of a generic bit b is

$$L(b) = \ln \frac{P(b=0|r)}{P(b=1|r)}. \tag{2}$$

Under the assumption of a-priori equally likely bits and taking into account the statistical distribution of the Gaussian noise, the LLR (2) can be expressed as

$$L(b) = \ln \frac{\sum_{s_0\in\mathcal{S}_0} e^{-|r-s_0|^2/\sigma^2}}{\sum_{s_1\in\mathcal{S}_1} e^{-|r-s_1|^2/\sigma^2}} \tag{3}$$

where $\mathcal{S}_0 \triangleq \{s\in\mathcal{S}|b=0\}$ and $\mathcal{S}_1 \triangleq \{s\in\mathcal{S}|b=1\}$. Observing that [10]

$$\sum_{s_\ell\in\mathcal{S}_\ell} e^{-|r-s_\ell|^2/\sigma^2} \simeq \exp\left(-\frac{1}{\sigma^2}\min_{s_\ell\in\mathcal{S}_\ell}|r-s_\ell|^2\right) \quad \ell=0,1$$

expression (3) of the LLR can be approximated as follows:

$$L(b) \simeq -\frac{1}{\sigma^2}\left(\min_{s_0\in\mathcal{S}_0}|r-s_0|^2 - \min_{s_1\in\mathcal{S}_1}|r-s_1|^2\right). \tag{4}$$

In the literature, the exact distribution of the bit LLRs at the output of a SO demodulator has been carefully investigated [11], [12]. In particular, it has been shown that at high[2] SNR $L(b)\sim\mathcal{N}(\mu_b\times(1-2b),\sigma_b)$, where μ_b (corresponding to the absolute value of the expectation) and σ_b depend on the specific QAM constellation [12].

As indicated in Fig. 1, we denote as $\{z_h\}$ the sequence of deinterleaved LLRs $\{L(b_h)\}$. Considering the use of a convolutional code, as in classical BICM schemes [3], [4], the branch metric used in the corresponding Viterbi decoder can be expressed as follows:

$$\begin{aligned}\lambda_j &= -\frac{|z_j-\mu_b\cdot(1-2c_j)|^2}{\sigma_b^2}\\ &\sim -[|z_j|^2 - 2\mu_b\cdot(1-2c_j)\cdot z_j + \mu_b^2\cdot\overbrace{|(1-2c_j)|^2}^{1}]\\ &\sim -|z_j|^2 - \mu_b^2 + 2\mu_b\cdot(1-2c_j)\cdot z_j\\ &\sim (1-2c_j)z_j\end{aligned} \tag{5}$$

where irrelevant (for the purpose of decoding) multiplicative and additive terms have been neglected. Taking into account the final branch metric expression (5), the multiplicative term $1/\sigma^2$ in the LLR approximate expression (4) for z_j is not relevant. Therefore, the following "simplified" LLR (sLLR) can be considered[3]:

$$L_{\mathrm{s}}(b_h) \triangleq \min_{s_1\in\mathcal{S}_1}|r-s_1|^2 - \min_{s_0\in\mathcal{S}_0}|r-s_0|^2. \tag{6}$$

Since in [13] it is shown that the metric (5) is the key ingredient also for soft-intput/soft-output decoders based on the forward-backward or sum-product algorithms, at the transmitter side of the scheme in Fig. 1 turbo or low-density parity-check codes could be used: the quantized demodulation approach proposed in Section III would still be applicable.

III. Quantized Soft-Output Demodulation of QAM Signals With Gray Labeling

A. Parametric Quantization

We preliminarily investigate the sLLR (6). In the presence of square QAM constellations with Gray labeling, a pair of (vertically or horizontally) adjacent QAM symbols differ by only one bit. For ease of exposition, we refer to an "inner" QAM symbol s with four adjacent symbols (this is typically the case for high-order QAM constellations) and we consider the symbol above it—the same considerations would hold considering the other adjacent symbols. Denote: s as the QAM symbol nearest to the received observable r (i.e., $s=s_{\mathrm{HO}}$ according to (1)); s' as the QAM symbol above s; b as the bit which changes from s to s'. Assuming to fix the origin of the coordinate system in s, i.e., $s=(0,0)$, it follows that $s'=(0,d)$. The corresponding geometric scenario is shown in Fig. 2 (a), where the coordinates of r are indicated as (x,y).

At this point, one can observe that, by construction: s is the symbol minimizing one of two minima at the right-hand side of the sLLR (6) relative to bit b; and s' is the symbol minimizing the other minimum. Defining $A\triangleq|r-s|$ and $B\triangleq|r-s'|$, one can write:

$$\begin{aligned}|L_{\mathrm{s}}(b)| &= |A^2-B^2|\\ &= |x^2+y^2-x^2-(y-d)^2| = |d(d-2y)|.\end{aligned} \tag{7}$$

Expression (7) indicates that the value of the sLLR depends only on y, i.e., it is constant over horizontal lines within the decision region of symbol s. Since $-d/2<y<d/2$, from (7) it follows:

$$|L_{\mathrm{s}}(b)| \le 2d^2 \tag{8}$$

where: the upper bound is reached when r falls on the lower border of the (square) decision region of s; and $|L_{\mathrm{s}}(b)|=0$ when r falls on the upper border of the decision region.

We now propose a simple parametric binary quantization approach. The extension to a more general (non-binary) quantization strategy is an interesting extension (as a multi-dimensional optimization has to be carried out). Denoting the quantized sLLR as $L_{\mathrm{q}}(b)$, its *sign* depends on bit b of symbol s (positive if $b=0$ and negative if $b=1$, according to the LLR

[2]We remark that at high SNR the absolute value of the LLR is high, i.e., the reliability is high. Therefore, according to the definition (2), the LLR is positive if $b=0$ and negative if $b=1$. This explains why the average value of the Gaussian distribution of the LLR can be written as $\mu_b\times(1-2b)$.

[3]Note that the sLLR in (6) is numerically different from the approximate LLR in (4).

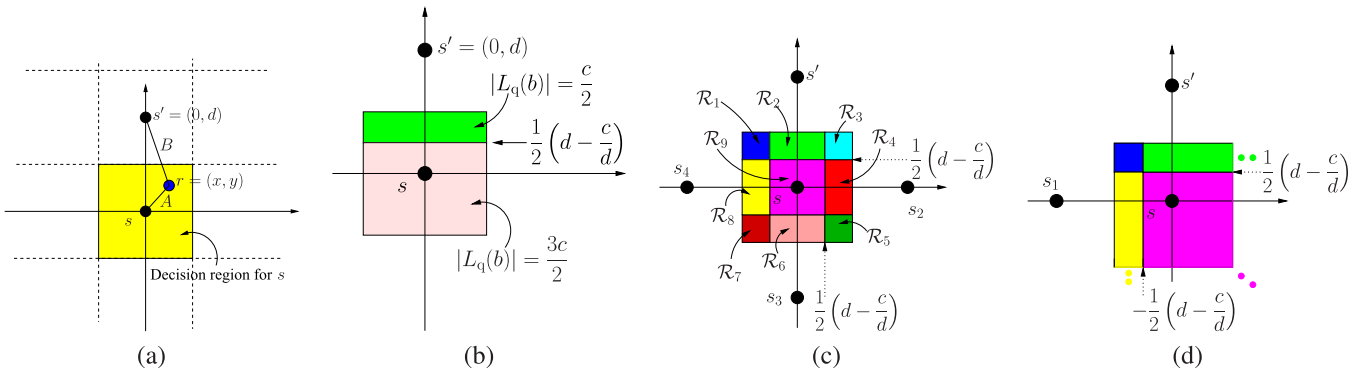

Fig. 2. Parametric binary quantization: (a) geometric representation for the evaluation of the sLLR of the bit b changing from s to s' (Gray labeling); (b) binary quantization regions for the sLLR of b; (c) the quantization sub-regions, relative to all bits, within the decision region relative to an inner symbol s; (d) the quantization sub-regions, relative to all bits, within the decision region relative to the symbol at the bottom right corner of the constellation.

definition (2)) and we set its *absolute value* as follows:

$$|L_q(b)| = \begin{cases} \dfrac{c}{2} & \text{if } |L_s(b)| \le c \\ & \iff \dfrac{1}{2}\left(d - \dfrac{c}{d}\right) \le y < \dfrac{d}{2} \\ \dfrac{3c}{2} & \text{if } |L_s(b)| > c \\ & \iff -\dfrac{d}{2} \le y < \dfrac{1}{2}\left(d - \dfrac{c}{d}\right) \end{cases} \tag{9}$$

where: $c \in (0, 2d^2)$ is a design real parameter, to be properly optimized; and $(d - c/d)/2$ can be interpreted as a per-dimension quantization threshold internal to the decision region. From a geometric point of view, the proposed two-level parametric optimization consists in dividing the decision region of s as shown in Fig. 2 (b) and selecting, depending on which sub-region the observable falls into, one of the two quantized values in (9) for the LLR. More precisely, in the upper sub-region the value of $|L_q(b)|$ would be between c (bottom border) and 0 (upper border) and we choose to associate all points with the arithmetic average $c/2$; similarly, in the lower region the value of $|L_q(b)|$ would be greater than c and we choose to associate all points with the larger value $3c/2$. This is the proposed geometric quantization strategy and, as will be discussed in Subsection III-D, allows to avoid direct computations of bit LLRs.

By generalizing the above approach, considering all 4 bits which change from an inner symbol s to the 4 adjacent constellation symbols, it is immediate to identify the 9 quantization sub-regions $\{\mathcal{R}_1, \ldots, \mathcal{R}_9\}$ shown in Fig. 2 (c). Depending on the quantization sub-region where r falls, it is possible to automatically identify the quantized values of the sLLRs of the 4 bits which change from s to the adjacent symbols. If the symbol s is not an inner symbol in the constellation, then the number of quantization sub-regions reduces: in Fig. 2 (d), we show the illustrative case when symbol s is the one at the bottom right corner of the constellation.

The above derivation (which leads to (9)) is valid for two-level quantization of the sLLR of a bit which changes from a symbol of interest $s = s_{HO}$ to one of its adjacent symbols (as shown above, the number of adjacent symbols can vary from 2 to 4). In the case of square constellations with Gray bit labeling and $M \ge 64$, i.e., when a symbol is associated with to more than 4 bits, regardless of the position of symbol s in the constellation at least 2 bits do not change from symbol s to any of its adjacent symbols. In this case, the absolute value of the sLLR of one of these bits should be set (according to a more refined quantization strategy) to a value higher than the upper value in (9). With a conservative approximation and for the sake of implementation simplicity, we still set the absolute values of the quantized sLLRs of these bits to $3c/2$.

B. Parametric Optimization

The quantization rule (9) depends on the parameter c, which can thus be optimized. The following optimization criterion is considered: for a specific coded QAM system and for a given value of the SNR, the optimized value of c, denoted as c_{opt}, is considered as the one associated with the lowest BER. The proposed simulation-based optimization needs to be applied offline (only once) per coded modulation scheme.

As an illustrative application example, we consider the scheme in Fig. 1 with a rate-3/4 convolutional code followed, after bit interleaving, by a 64-QAM or a 256-QAM modulator. In particular, the rate-3/4 convolutional code is obtained from the rate-1/2 convolutional code, with generators $(171)_8 = (001111001)_2$ and $(133)_8 = (001011011)_2$ [8], punctured with the following pattern:

$$PP_{3/4} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

The considered QAM constellation is formed by points $\{\pm i, \pm j\}$, with $i, j \in \{1, 3, 5, 7\}$, so that $d = 2$. In Fig. 3, the BER is shown, as a function of c, considering both 64-QAM (dashed lines) and 256-QAM (solid lines) for various values of the bit SNR E_b/N_0, where E_b is the average energy per bit and $N_0 = \sigma^2$. From the obtained results, it emerges clearly that, *regardless of the SNR value*, there exists an optmized value c_{opt} for each coded modulation scheme: 2.2 for 64-QAM and 3.0 for 256-QAM. This suggests that the optimized subregion mappings shown in Fig. 2 (c) and Fig. 2 (d) depends only on the coded modulation scheme. A rigorous proof of this conjecture is an open problem.

C. BER Performance

We now investigate the performance, in tems of BER as a function of the bit SNR, with various demodulation strategies, considering both 64-QAM and 256-QAM BICM schemes. In Fig. 4, the following systems are considered: (i) BICM with HO demodulation (curves with squares);

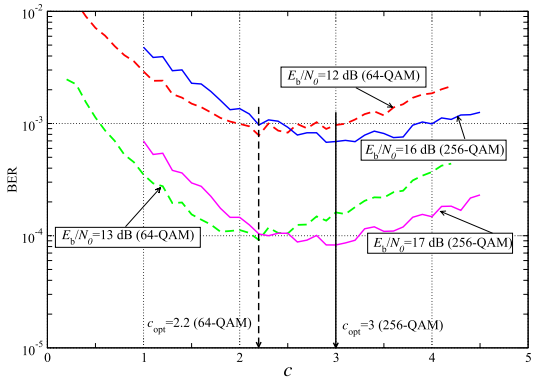

Fig. 3. BER, as a function of c, with a BICM with a rate-3/4 convolutional code mapped over a 64-QAM (dashed lines) or 256-QAM (solid lines). Various values of the SNR are considered and the optimal values of c are indicated.

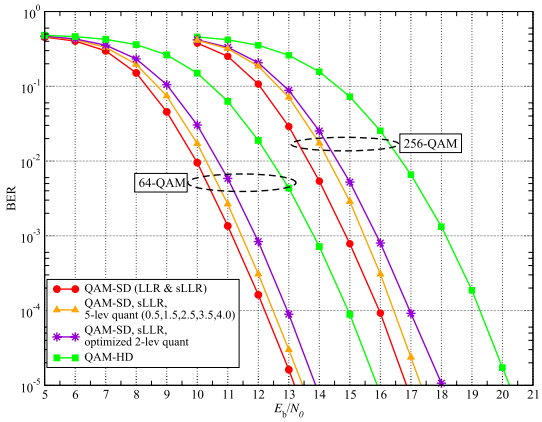

Fig. 4. BER, as of function of the SNR, with a rate-3/4 punctured CC(171,133) mapped onto a 64-QAM or 256-QAM, with various demodulation strategies (HO, unquantized SO, SO with 2-level quantization, SO with 5-level quantization).

(ii) BICM with two-level quantized SO demodulation, setting c to c_{opt} from Fig. 3, namely to: 2.2 for 64-QAM and 3.0 for 256-QAM (curves with stars); (iii) BICM with unquantized SO demodulation (curve with circles). It can be observed that the two-level quantized SO demodulation approach, with internal per-dimension quantization threshold $(d - c/d)/2$ (equal to 0.45 and 0.25 for 64-QAM and 256-QAM, respectively) and absolute values of the quantized sLLRs given by (9) (namely, $\{1.1, 3.3\}$ for 64-QAM and $\{1.5, 4.5\}$ for 256-QAM), incurs a limited performance loss (0.7 dB and 1 dB for 64-QAM and 256-QAM, respectively, at a BER equal to 10^{-4}) with respect to that with unquantized SO demodulation. For comparison purposes, we also show the results considering a heuristic five-level quantized SO demodulation, with quantized sLLR values $\{0.5, 1.5, 2.5, 3.5, 4.0\}$ (curves with triangles): the improvement, with respect to the two-level quantized case, is very limited (smaller than 0.5 dB for both 64-QAM and 256-QAM).

D. Implementation Complexity

The proposed two-level quantized SO demodulation strategy lends itself to an attractive look-up table-based implementation of a quantized SO demodulator, avoiding the computation of bit-level LLRs followed by quantization. Considering an "inner" symbol with 4 adjacent symbols (Fig. 2 (c)), it is sufficient to identify the quantization sub-region where the observable r falls—recall that the number of quantization sub-regions reduces for symbols on the constellation borders (as shown in Fig. 2 (d)). For each quantization sub-region (there are at most 9 sub-regions), one needs 1 bit to indicate the value of the absolute value of the quantized sLLR, i.e., $c/2$ or $3c/2$. In the case of M-ary square constellation, for each decision subregion one needs a $\log_2 M$ array (associated with each bit), where each entry contains 1 bit to indicate the value of the quantized sLLR. The memory occupation (in terms of bits) required by the look-up table for the proposed quantization strategy is thus at most $9 \times M \times \log_2 M$ i.e., $O(M \times \log_2 M)$.

IV. Conclusions

In this paper, we have proposed a low-complexity parametric approach, based on a geometric intepretation, to two-level quantization of bit LLRs at the output of a SO demodulator for QAM signals with Gray labeling. The application of the proposed approach to a coded QAM system with separate demodulation and decoding (namely, an illustrative BICM scheme) has been considered. The obtained results show that optimized two-level quantization entails a limited performance loss with respect to "classical" unquantized SO demodulation. The proposed approach allows an efficient implementation of a look-up table-based quantized SO demodulator, making it very attractive from a practical viewpoint. In order to investigate the "optimality" of our geometric approach, an interesting research direction consists in comparing it with a quantization strategy based on the maximization of the mutual information between observables and transmitted symbols (with verification by capacity-approaching codes).

References

[1] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[2] M. Seimetz, *High-Order Modulation for Optical Fiber Transmission*, vol. 143. Berlin, Germany: Springer, 2009.
[3] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 873–884, May 1992.
[4] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.
[5] M. Baldi, F. Chiaraluce, and G. Cancellieri, "Finite-precision analysis of demappers and decoders for LDPC-coded M-QAM systems," *IEEE Trans. Broadcast.*, vol. 55, no. 2, pp. 239–250, Jun. 2009.
[6] C. Novak, P. Fertl, and G. Matz, "Quantization for soft-output demodulators in bit-interleaved coded modulation systems," in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, South Korea, Jun./Jul. 2009, pp. 1070–1074.
[7] C. J. R. Runge and J. R. C. Piqueira, "A geometric approach for turbo decoding of Reed-Solomon codes in QAM modulation schemes," *IEEE Commun. Lett.*, vol. 13, no. 5, pp. 333–335, May 2009.
[8] J. G. Proakis, *Digital Communications*, 4th ed. New York, NY, USA: McGraw-Hill, 2001.
[9] F. Aurenhammer, "Voronoi diagrams—A survey of a fundamental geometric data structure," *ACM Comput. Surv.*, vol. 23, no. 3, pp. 345–405, 1991.
[10] A. J. Viterbi, "An intuitive justification and a simplified implementation of the MAP decoder for convolutional codes," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 260–264, Feb. 1998.
[11] A. Alvarado, L. Szczecinski, R. Feick, and L. Ahumada, "Distribution of L-values in Gray-mapped M^2-QAM: Closed-form approximations and applications," *IEEE Trans. Commun.*, vol. 57, no. 7, pp. 2071–2079, Jul. 2009.
[12] A. Kenarsari-Anhari and L. Lampe, "An analytical approach for performance evaluation of BICM transmission over Nakagami-*m* fading channels," *IEEE Trans. Commun.*, vol. 58, no. 4, pp. 1090–1101, Apr. 2010.
[13] G. Ferrari, G. Colavolpe, and R. Raheli, "A unified framework for finite-memory detection," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 9, pp. 1697–1706, Sep. 2005.