Optimum Power Allocation and Bit Loading for BICM Systems

David Matas and Meritxell Lamarca

***Abstract*—This paper introduces a joint bit loading and power allocation algorithm for systems combining bit-interleaved coded modulation (BICM) with multicarrier transmission. The proposed algorithm maximizes the mutual information, so it can be regarded as a generalization of mercury/waterfilling policy that incorporates bit loading.**

The followed approach relies on irregular modulation and power to cast the problem in the framework of convex optimization. This allows to derive the optimum solution without resorting to greedy algorithms, embedding the bit loading in the definition of an equivalent constellation such that the complexity increase with respect to mercury/waterfilling is negligible.

While irregular modulation plays a key role in algorithm definition, it is proved that only a few subcarriers employ it and it is shown that a practical low complexity algorithm can be obtained with minimal losses that does not use irregular modulation.

***Index Terms*—Power allocation, bit loading, adaptive modulation, bit-interleaved coded modulation (BICM), channel state information (CSI), mutual information, OFDM.**

I. Introduction

The combination of bit-interleaved coded modulation (BICM) [1] with orthogonal frequency division multiplexing (OFDM) [2] provides a low-complexity nearly-optimal performance approach to the broadband transmission in multipath scenarios. On the one hand, the paradigm of channel code and modulation separation by means of a bit interleaver introduced by BICM has been proved a versatile approach to spectrally efficient transmission in fading channels. BICM allows a large flexibility in constellation selection and channel coding design at the expense of minor performance losses. On the other hand, the use of OFDM allows to get rid of intersymbol interference converting the frequency selective channel into a set of parallel non-interfering channels thanks to the use of multicarrier transmission. The good performance vs. complexity trade-off provided by BICM-OFDM has motivated a broad interest in this transmission scheme, as well as its introduction in many standards (e.g. IEEE802.11a/g, IEEE802.16).

This paper deals with transmission schemes for BICM-OFDM systems subject to slow-fading. In low-mobility scenarios the channel state can be accurately tracked by both the transmitter and the receiver and the performance can be improved adapting the signaling to the instantaneous channel spectral shape. In the BICM-OFDM scheme, nearly optimal performance can be achieved by bit loading and power allocation, i.e. assigning a different constellation (number of bits) and different power to each subcarrier and each channel access according to the channel frequency response.

Probably the most well-known criterion for power allocation is the maximization of the channel capacity under an average power constraint. The solution for the additive white Gaussian noise (AWGN) channel is known as waterfilling [3, Sec 10.4]. Unfortunately, this solution assumes a continuous and Gaussian modulation so it is not optimum for practical systems using discrete constellations, since the bit rate assignments are constrained to be integer. In this case, if the bit loading has been previously determined, the power allocation that maximizes the mutual information is provided by the mercury/waterfilling policy (MWF) [4].

Taking into account that state of the art channel codes can achieve performance very close to the Shannon limit, the mutual information provides a good criterion for system design without taking into account the specific channel code employed. However, since MWF only optimizes the power allocation, a bit loading algorithm is also required. Unfortunately, mutual information cannot be employed as the optimization criterion for bit loading design when typical constellations (e.g. m-QAM, m-PSK) and optimum transceivers are considered, since for a fixed SNR the largest constellation always provides the largest mutual information. In other words, the mutual information is maximized employing the largest constellation available at any SNR. Therefore, other methods have been proposed in the literature that introduce practical constraints related to the bit error rate (BER) performance, either for the uncoded or coded case [5]–[10]. However, this situation may change if we take account of the mutual information loss introduced by suboptimum transceivers. In the case of BICM, the magnitude of this loss depends on the constellation, its labeling and the SNR. As it is well known, for QAM constellations with Gray labeling the MI of small constellations surpasses that one of larger ones at low SNR's (see [1, Fig. 6]). Hence, when BICM schemes are considered, mutual information provides a meaningful criterion for joint optimization of bit loading and power

Paper approved by X. Wang, the Editor for Multiuser Detection and Equalization of the IEEE Communications Society. Manuscript received March 20, 2008; revised June 2, 2009 and November 16, 2009.

The authors are with the Department of signal Theory and Communications, Universitat Politècnica de Catalunya (UPC), Campus Nord, D5, c/Jordi Girona 1-3, 08034 Barcelona, Spain. Since May 2009 M. Lamarca is on leave of absence at the Electrical and Computer Engineering Department, University of Delaware, Newark, DE 19716 USA (e-mail: dmatas@gps.tsc.upc.edu, lamarca@udel.edu).

Part of this work has been presented at the 2007 IEEE Global Communications Conference, Globecom 2007, Washington DC, November 2007. This work has been partially funded by the Spanish/Catalan Science and Technology Commissions and by FEDER funds from the European Commission: TEC2007-68094-C02-02, 2005SGR-00639

Digital Object Identifier 10.1109/TCOMM.2010.062310.080128

allocation that can be applied to design transceivers with nearly optimal performance.

In this paper we propose a bit loading and power allocation algorithm for BICM-OFDM systems that maximizes the mutual information under an average power constraint. We rely on irregular modulation to formulate the problem in the framework of a convex optimization and to derive a low complexity algorithm that does not require the use of iterative numerical procedures nor greedy algorithms. The proposed algorithm does not provide significant gains in terms of spectral efficiency increase, but rather provides a new tool for joint bit loading and power allocation that maximizes the mutual information while it maintains the computational complexity of mercury/waterfilling.

Although the algorithm is proposed for a BICM-OFDM transmission, it could be employed in any other scenario where parallel subchannels arise (e.g. multiple antenna transmission where the MIMO channel is diagonalized through the use of linear pre/post-filtering based on its singular value decomposition) and where the transmission scheme results in mutual information vs. SNR curves that overlap for different configurations (e.g. APSK constellations).

The rest of the paper is organized as follows. In section II we describe the system model employed in section III to formulate the bit loading and power allocation policy. The solution is divided in two steps, both expressed as a convex optimization problem, solved and analyzed in sections IV and V respectively. Concluding remarks are provided in section VI.

II. Irregular Modulation and Power Approach

In this section we describe the system set-up in which the proposed bit loading and power allocation policy will be employed. It consists on the application of irregular modulation and power allocation to a BICM scheme in an scenario with multiple parallel subchannels [1].

Irregular modulation was first proposed by Schreckenbach and Bauch [11]. It combines symbols belonging to different constellations within the same transmission block, despite the channel state remaining constant. This transmission scheme is described by the amount of time in which bits are mapped to each one of the constellations. While the application of conventional adaptive coding and modulation to the Gaussian channel results in a step-wise throughput vs. SNR curve, the use of irregular modulation provides a smooth curve without the need for a fine code rate granularity [11]. This idea has been extended to irregular modulation with irregular power [12], [13] for the design of practical bit loading schemes that optimize the performance taking into account the channel coding stage. This is also the approach employed in this paper.

Figure 1 depicts the system block diagram, including the coding, interleaving, mapping and power allocation stages. Following a BICM scheme, the coded bits are bit-interleaved and delivered to the modulation and power allocation stages. Consider Q parallel subchannels (subcarriers in the case of OFDM) with coefficients $\{H_1, ..., H_Q\}$. Denote by $p_q(n)$ the power allocated to and by $x_q(n)$ the unit-power symbol transmitted through the q-th subchannel in the n-th channel access. The symbol belongs to one of the N available constellations, $\{\mathcal{C}_1, ..., \mathcal{C}_N\}$, with $\{m_1, \cdots, m_N\}$ bits per symbol respectively and possibly different labeling [2]. Then, the received signal is

$$y_q(n) = H_q\sqrt{p_q(n)}x_q(n) + w_q(n) \quad q = 1, ..., Q \tag{1}$$

where $w_q(n)$ is the additive complex white Gaussian noise term of zero mean and variance σ^2, independent among subchannels. At the receiver, the suboptimum detector computes the bit log-likelihood ratios (LLR's) of the transmitted bits according to the procedure described in the Appendix B, deinterleaves and delivers them to the decoder.

According to the irregular modulation and power scheme, we allow the transmission of symbols belonging to different constellations with different allocated power within the same subchannel, and we let this configuration to be different for each subchannel. If more than one constellation is used within a subchannel, the order in which the bits are mapped to them is predefined and, therefore, known at both transmitter and receiver. Let α_{iq} be the fraction of symbols transmitted through the q-th subchannel that belong to constellation $\mathcal{C}_i$, and let p_{iq} be the power allocated to each one of them. According to their definition, these parameters must fulfill:

$$\alpha_{iq} \geq 0, \quad p_{iq} \geq 0 \quad and \quad \sum_{i=1}^{N} \alpha_{iq} \leq 1 \tag{2}$$

for $q = 1, ..., Q$ and $i = 1, ..., N$. If $\alpha_{iq} = 0$, then constellation $\mathcal{C}_i$ is not employed in the q-th channel. If $\sum_{i=1}^{N} \alpha_{iq} = 0$ then the q-th subchannel is not used and if $0 < \sum_{i=1}^{N} \alpha_{iq} < 1$ then it is used during a fraction of the channel accesses. Note that this formulation can be mathematically seen as a continuous relaxation of the usual bit allocation, in which $\alpha_{iq} \in \{0, 1\}$ (i.e., only one constellation can be used per subchannel).

III. Problem Formulation

The objective of this paper is the derivation and analysis of a bit loading and power allocation algorithm that maximizes the mutual information for the system set-up defined in section II.

Using the parameters previously defined, the power allocated to and the MI of the q-th subchannel are obtained as the weighted averages of the individual values for each constellation,

$$P_q = \sum_{i=1}^{N} \alpha_{iq} p_{iq} \tag{3}$$

$$\bar{I}_q = \sum_{i=1}^{N} \alpha_{iq} I_i(p_{iq}\gamma_q) \tag{4}$$

where $\gamma_q = |H_q|^2/\sigma^2$ is a measure of the subchannel reliability (the SNR with unit transmitted power) and $I_i(\mu) =$

[1] In the sequel we denote the OFDM subcarriers as subchannels to emphasize that the proposed algorithm could also be applied to any system model that could be described by parallel channels.

[2] Note that BICM performance in terms of MI is sensitive to constellation labeling.

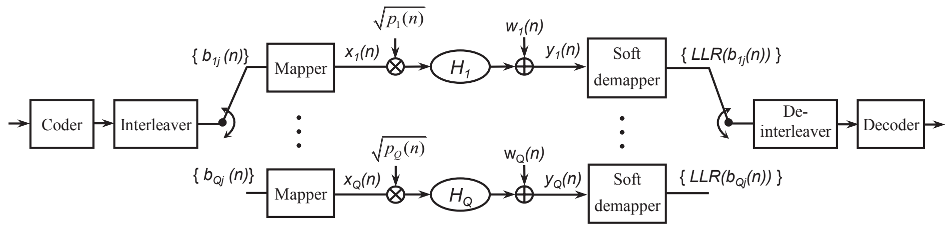


Fig. 1. Block diagram of the BICM scheme.

$I_i\left(x; LLR(b_1), ..., LLR(b_{m_i})\right)$ is the MI between the transmitted symbols and the corresponding bit LLR's at the demodulator output for the i-th constellation when it is employed in an AWGN channel with SNR μ.

Finally, the total MI averaged over all subchannels is

$$\bar{I} = \frac{1}{Q}\sum_{q=1}^{Q}\bar{I}_q \tag{5}$$

and the average constellation size is

$$\eta = \frac{1}{Q}\sum_i\sum_q \alpha_{iq} m_i \tag{6}$$

The optimum bit loading and power allocation is defined as the one that maximizes the MI in (5) with respect to α_{iq} and p_{iq} subject to the constraints in (2) and the average power constraint, that is,

$$\max_{\{p_{iq}\},\{\alpha_{iq}\}} \frac{1}{Q}\sum_{q=1}^{Q}\sum_{i=1}^{N}\alpha_{iq} I_i\left(p_{iq}\gamma_q\right) \tag{7a}$$

$$s.t. \quad \alpha_{iq} \geq 0\,, \quad p_{iq} \geq 0\,, \quad \begin{matrix} i = 1, ..., N \\ q = 1, ..., Q \end{matrix} \tag{7b}$$

$$\sum_{i=1}^{N}\alpha_{iq} \leq 1 \qquad q = 1, ..., Q \tag{7c}$$

$$\frac{1}{Q}\sum_{q=1}^{Q}\sum_{i=1}^{N}\alpha_{iq} p_{iq} \leq P_T \tag{7d}$$

where P_T is the maximum available power at the transmitter.

Introducing the average power allocated per subchannel defined in (3) into these equations, one can observe that the joint optimization of the parameters for all subchannels can be formulated as a two step optimization.

The first step is the optimization of the power allocation and bit loading for a single AWGN channel. Let us consider a generic AWGN channel with reliability γ and allocated power P. In this case, we drop the subindex q indicating the subchannel and, therefore, we denote the parameters as α_i and p_i instead of α_{iq} and p_{iq}. If we introduce the normalized power parameters $p_i' = p_i/P$, then the first optimization can be expressed as a function of the SNR $\mu = P\gamma$ as

$$\bar{I}_o(\mu) = \max_{\{p_i'\},\{\alpha_i\}} \sum_{i=1}^{N}\alpha_i I_i\left(p_i'\mu\right) \tag{8a}$$

$$s.t. \quad \textstyle\sum_{i=1}^{N}\alpha_i p_i' = 1 \tag{8b}$$

together with the inequality constraints (7b) and (7c).

The second step consists in the following problem of power allocation over parallel subchannels:

$$\max_{\{P_q\}} \frac{1}{Q}\sum_{q=1}^{Q}\bar{I}_o(\gamma_q P_q) \tag{9a}$$

$$s.t. \quad \textstyle\frac{1}{Q}\sum_{q=1}^{Q} P_q \leq P_T \tag{9b}$$

Hence, in the second step no bit loading must be done. As we explain in more detail in section V, this power allocation is solved by mercury/waterfilling policy considering a single irregular modulation resulting from the first step.

These two steps are analyzed in the following two sections. The single AWGN case is studied in section IV (first step). Afterwards, section V generalizes the design for multiple parallel channels (second step).

IV. Single AWGN Channel

This section presents the solution to the single channel optimization expressed in equation (8), corresponding to the first step of the global optimization. The problem in equation (8) is not convex in terms of (α_i, p_i'). However, as the MI functions are concave, introducing the change of variables $s_i = p_i'\alpha_i$ it can be expressed it as the following convex optimization problem:

$$\bar{I}_o(\mu) = \max_{\{s_i\},\{\alpha_i\}} \sum_{i=1}^{N}\alpha_i I_i\left(\frac{s_i}{\alpha_i}\mu\right) \tag{10a}$$

$$s.t. \quad \alpha_i \geq 0\,, \quad s_i \geq 0\,, \quad i = 1, ..., N \tag{10b}$$

$$\sum_{i=1}^{N} s_i = 1 \tag{10c}$$

$$\sum_{i=1}^{N}\alpha_i \leq 1 \tag{10d}$$

A. Solution

The optimization problem in (10) can be regarded as a multiple objective optimization based on a weighted sum of the set of objective functions, where the weights are not constant parameters but also variables to be optimized. For our specific setting, in which each one of these functions (the mutual information of each constellation) depends on only one of the variables (the power allocated to it), the solution is their envelope or convex hull, determined by the tangency line among all the individual functions. Next we provide the final expression of the solution, while its derivation can be found in Appendix A.

Let us first introduce some definitions:

- Derivative of the mutual information with respect to SNR for constellation $\mathcal{C}_i$:

$$D_i(\mu) = \frac{dI_i(\mu)}{d\mu} \tag{11}$$

- Legendre transform [14] of the mutual information:

$$L_i(x) = C_i\left(D_i^{-1}(x)\right) \tag{12}$$

where $C_i(\mu)$ is the gap between the MI and its first order Taylor approximation at the origin ($\mu = 0$):

$$C_i(\mu) = I_i(\mu) - \rho D_i(\mu) \tag{13}$$

In order to find a closed expression for this solution let us define a partition of the domain of the Legendre transforms of the MI functions into the following regions:

$$\mathcal{R}_i = \{x \quad : \quad L_i(x) \geq L_j(x) \quad \forall j \neq i\} \tag{14}$$

If one of these regions is empty, then the corresponding constellation is never used and we assume it to be removed from the set of possible ones. This condition is equivalent to the MI of the corresponding constellation being smaller than the MI provided by other constellations at all SNR's. Let us assume that the constellations of this set are indexed according to the order of appearance of these regions for a decreasing x, i.e.

$$x_i \in \mathcal{R}_i \quad and \quad x_{i+1} \in \mathcal{R}_{i+1} \quad \Rightarrow \quad x_i \geq x_{i+1}$$

Besides, let us define ν_i as the boundary points between pairs of contiguous regions, i.e.

$$\nu_i = x \quad : \quad x \in \mathcal{R}_i \quad and \quad x \in \mathcal{R}_{i+1} \quad for \quad i = 1, ..., N-1$$

or, in other terms, as the points in which the Legendre transforms of the MI of the corresponding constellations intersect

$$L_i(\nu_i) = L_{i+1}(\nu_i) \quad for \quad i = 1, ..., N-1 \tag{15}$$

Since there are N constellations whose region $\mathcal{R}_i$ is not empty, there are $N-1$ boundary points.

Figure 2 depicts function $L_i(x)$ for different constellations. The regions defined in equation (14) and the intersections in (15) are also indicated. In geometric terms, when two Legendre transforms intersect the tangent to the corresponding MI functions coincide. These points determine the range of SNR's in which the corresponding constellations are combined, which are bounded by the following thresholds

$$\begin{array}{ll} \mu_i^- = D_i^{-1}(\nu_{i-1}) & for \quad i = 2, ..., N \\ \mu_i^+ = D_i^{-1}(\nu_i) & for \quad i = 1, ..., N-1 \end{array} \tag{16}$$

together with $\mu_0^+ = 0$, $\mu_1^- = 0$, $\mu_N^+ = \infty$ and $\mu_{N+1}^- = \infty$.

As derived in Appendix A, the solution of the optimization problem (10) can be expressed in terms of these thresholds as a function of the SNR as follows:

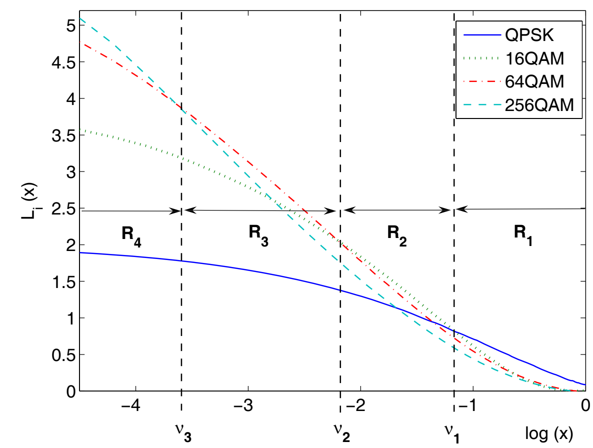


Fig. 2. Legendre transform of the mutual information for different constellations for the BICM scheme described in Appendix B. The subindexes are assigned in the following order: 1:QPSK, 2:16-QAM, 3;64-QAM, 4:256-QAM. The regions defined in equation (14) are depicted, as well as the boundary points ν_i defined in (15).

- The fraction of use of each constellation is defined piece-wise by linear sections:

$$\alpha_i(\mu) = \begin{cases} 0 & if \quad 0 \leq \mu \leq \mu_{i-1}^+ \\ k_i(\mu - \mu_{i-1}^+) & if \quad \mu_{i-1}^+ < \mu < \mu_i^- \\ 1 & if \quad \mu_i^- \leq \mu \leq \mu_i^+ \\ k_{i+1}(\mu_{i+1}^- - \mu) & if \quad \mu_i^+ < \mu < \mu_{i+1}^- \\ 0 & if \quad \mu_{i+1}^- \leq \mu < \infty \end{cases} \quad i = 1, ..., N \tag{17}$$

with

$$k_j = \frac{1}{\mu_j^- - \mu_{j-1}^+} \quad j = 1, ..., N+1$$

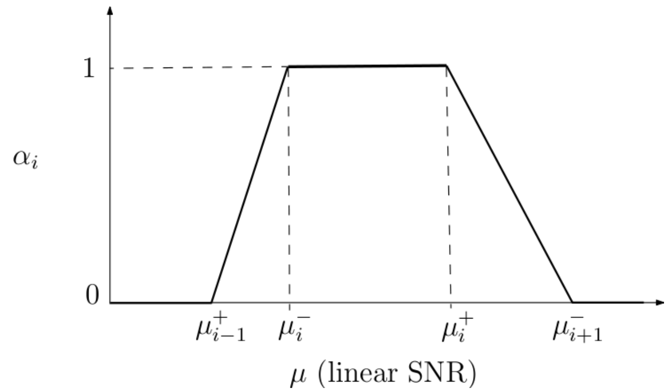
- The normalized power allocated to each constellation is

$$p_i'(\mu) = \begin{cases} 0 & if \quad 0 \leq \mu \leq \mu_{i-1}^+ \\ \frac{\mu_i^-}{\mu} & if \quad \mu_{i-1}^+ < \mu < \mu_i^- \\ 1 & if \quad \mu_i^- \leq \mu \leq \mu_i^+ \\ \frac{\mu_i^+}{\mu} & if \quad \mu_i^+ < \mu < \mu_{i+1}^- \\ 0 & if \quad \mu_{i+1}^- \leq \mu < \infty \end{cases} \quad i = 1, ..., N \tag{18}$$

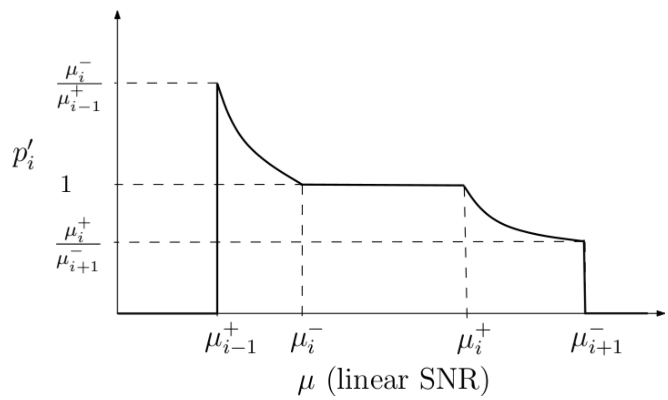
Figures 3(a) and 3(b) depict these functions. Combining equations (17) and (18) we can easily find the expression of the optimum mutual information as a piece-wise function:

$$\bar{I}_o(\mu) = \begin{cases} I_{i-1}(\mu_{i-1}^+) + \frac{I_i(\mu_i^-) - I_{i-1}(\mu_{i-1}^+)}{\mu_i^- - \mu_{i-1}^+}(\mu - \mu_{i-1}^+), & \mu_{i-1}^+ < \mu < \mu_i^- \\ I_i(\mu) & , \mu_i^- \leq \mu \leq \mu_i^+ \end{cases} \quad i = 1, ..., N \tag{19}$$

for $i = 1, ..., N$ and where we have defined $I_0(\mu) = 0 \quad \forall \mu$. Notice that the mutual information is linear with SNR in the SNR values where two constellations are combined. For a generic AWGN channel with reliability γ and with maximum transmitted power P, the solution is found applying the previous equations for an SNR $\mu = \gamma P$.



(a) α_i



(b) p'_i

Fig. 3. Graphical representation of the solution for the i-th constellation: fraction of use (a) and allocated powers (b).

B. Analytical and graphical interpretation

The resulting irregular modulation and power allocation expressed in equations (17) and (18) can be seen as an equivalent constellation. Given a set of available constellations, this equivalent constellation is obtained as the SNR-dependent constellation selection that provides the maximum MI for each SNR and is characterized by the SNR vs. MI curve in equation (19). In this section we illustrate the use of irregular modulation and power for MI maximization in the AWGN channel showing its application to a BICM scheme with unit available power and QPSK, 16QAM, 64QAM and 256QAM with Gray labeling as possible constellations.

Figure 4 represents the resulting optimum MI vs. the SNR compared to the ones of the individual constellations. As evidenced in the zoomed region, in the SNR intervals where two constellations are combined, the optimum mutual information curve as a function of the SNR is linear (and then, the first derivative is constant) and it coincides with the common tangent to the MI curves of these two constellations. This property determines the gain achieved with the irregular modulation: it corresponds to the gap between the MI curves of the constellations and their tangent. While this gain might be considered to be small, the fact that the optimum MI is concave with respect to the SNR is of paramount importance to derive a bit loading and power allocation algorithm for multiple parallel channels.

The solution in terms of parameters α_i and p'_i for each constellation has the shape depicted in Figs. 3(a) and 3(b)

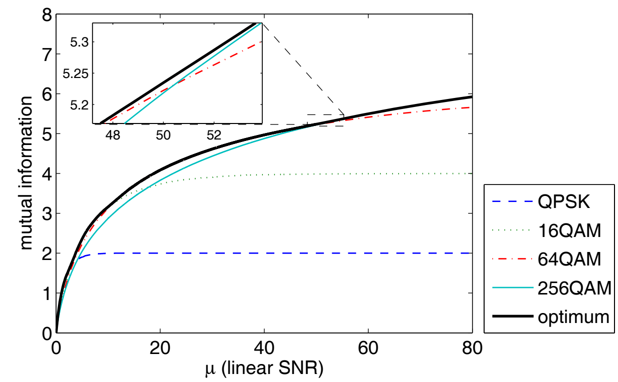


Fig. 4. Optimum mutual information ($\bar{I}_o(\mu)$) for the AWGN channel example. The individual MI curves are also shown.

respectively. In the intervals where an irregular modulation is employed, the α_i of the two constellations that are combined have opposite behaviors, one decreasing and the other one increasing, with a constant slope of the same absolute value. The normalized power p'_i is inversely proportional to the SNR for both, so the mutual information per constellation symbol remains constant and equal to the value in which the tangent line intersects with the MI curves.

Finally, Fig. 5 presents the mutual information per bit (i.e., $\bar{I}_o/\eta$) that would be obtained in the proposed example. This mutual information dictates the rate of the binary channel code that would be incorporated in the BICM scheme. As can be observed, the MI per bit keeps within the range $[0.5, 0.83]$ in the SNR interval of 0 to 20 dB, indicating that it would be easy to design capacity-approaching binary codes operating at these rates. Note that if bit loading was not applied and the largest constellation was always used the rate of required binary channel code would be very small at low SNR's and, hence, very difficult to design.

In Fig. 5, it can also be appreciated the piecewise behavior of the solution: the MI per bit coincides with the one of every constellation in a in certain region, corresponding to the SNR values in which only that constellation is employed. To make this clear, the average size of the resulting optimum constellation (defined in equation (6),with $Q = 1$) has been plotted also in the same figure. The regions in which the solution selects a single constellation are characterized by a constant value of the constellation size, whereas an irregular combination of two constellations in associated to the constellation size transitions.

V. Multiple Parallel Channels

A. Solution

The second optimization step, expressed in equation (9), can be regarded as a power allocation problem over Q parallel subchannels, all of them using the same equivalent constellation that provides a mutual information $\bar{I}_o(\mu)$. Although mercury/waterfilling (MWF) was originally formulated for linear modulations and lossless transceivers, it can be easily shown that it can be applied also to allocate power in other transmission schemes in parallel subchannels as long as the mutual information is concave with the SNR and we replace

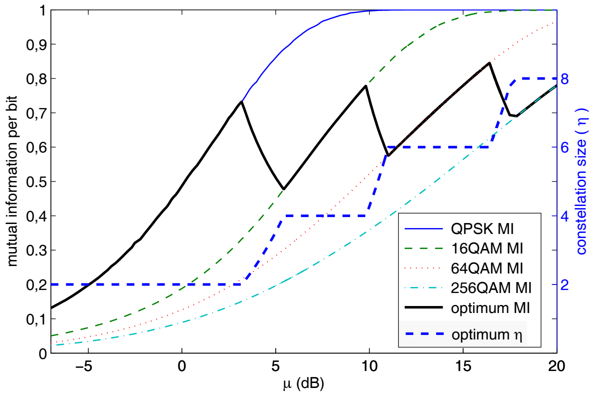


Fig. 5. Mutual information per bit (I_o/η) and average constellation size (η) for the optimum design in a single AWGN channel (left and right axis respectively). The mutual information curves for all available constellations (I_i/m_i) are also shown.

the mean square error in the formulation in [4] by the MI derivative with respect to the SNR. As the MI for each constellation I_i is concave with the SNR, the single channel optimum $\bar{I}_o$ presented in the previous section is also concave. Therefore, defining the derivative of this function as

$$D_o(\mu) = \frac{d\bar{I}_o(\mu)}{d\mu} \tag{20}$$

the solution to (9) is provided by the MWF expression:

$$P_q = \begin{cases} \frac{1}{\gamma_q} D_o^{-1}\left(\frac{\delta}{\gamma_q}\right) & if \quad \delta \leq \gamma_q \cdot D_o(0) \\ 0 & otherwise \end{cases} \tag{21}$$

where the waterlevel δ is selected in order to fulfill the power constraint.

B. Analytical and graphical interpretation

The procedure to allocate power and its graphical interpretation presented in [4] apply here, since the optimum solution is given by mercury/waterfilling. However, it is interesting to note which are the implications of the particular shape of the single channel optimized MI.

Following the procedure described in [4], the waterlevel δ can be found depicting the scaled derivative function of the mutual information for each subchannel as a function of the allocated power, and looking for the waterlevel δ such that the corresponding abscissas satisfy the average power constraint. Figure 6 illustrates this procedure for a toy example with $Q = 4$, $P_T = 1$ and the same constellation parameters employed in the example in section IV-B (which resulted in the equivalent constellation whose MI vs. SNR plot is depicted in Fig. 4). We consider the gains H^2 of the four parallel subchannels equal to $\frac{1}{24}[12, 6, 5, 1]$ and an SNR of 13dB (i.e. $\gamma_1 = 9.98$, $\gamma_2 = 4.99$, $\gamma_3 = 4.16$ and $\gamma_4 = 0.83$). Figure 6 depicts the scaled function $D_o(\mu)$ (derivative of the MI plot in Fig. 4) for each subchannel. As shown in the figure, the waterlevel is $\delta = 0.965$ and the optimum power values are $P_1 = 1.43$, $P_2 = 1.138$, $P_3 = 1.1438$ and $P_4 = 0.29$. Hence, from equations (17) and (18), the optimum bit loading variables are ($\alpha_{31} = \alpha_{22} = \alpha_{23} = \alpha_{14} = 1$), ($p'_{31} = p'_{22} = p'_{23} = p'_{14} = 1$), being null the remaining ones.

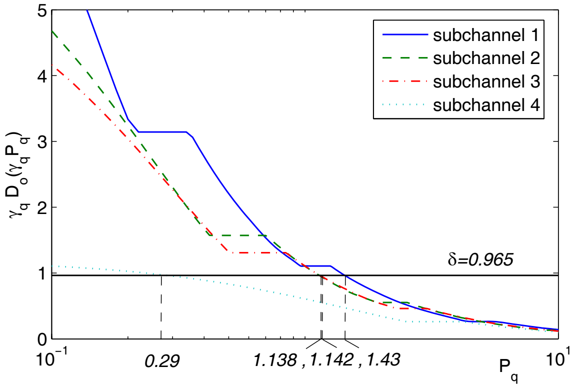


Fig. 6. Weighted derivatives chart representing the MWF solution.

As seen in Fig. 6, and described in section IV-B, the derivative of the single channel optimum MI, $D_o(\mu)$, is constant in the range of SNR's where two constellations are combined. Hence, the inverse of the derivative is not well defined in these sections. If the optimum δ corresponds to a flat section of $\gamma_q D_o$ for only one of the subchannels, then the ambiguity in the power value for that subchannel can be solved from the knowledge of the power for the other subchannels and the application of the average power constraint. However, if the γ_q values are such that the optimum δ coincides with a flat section for several subchannels, then the optimization problem has multiple solutions, all of them providing the same optimum mutual information value. In that case, an additional criterion should be introduced to select one of the solutions.

Let us consider the transmission using BICM through 32 parallel subchannels with gains depicted in Fig. 7 employing the same set of constellations as the AWGN channel example (QPSK, 16QAM, 64QAM and 256QAM with Gray labeling). Figure 8 depicts the optimum MI vs. SNR provided by the bit loading and power allocation algorithm proposed in this paper for this scenario and compares it with the capacity obtained by classic waterfilling. Note that the difference between them is due to the limitation in the constellation size and the lack of shaping gain. For completeness, the mean constellation size is also depicted.

It is important to remark that, as the maximum number of boundary points of the regions $\mathcal{R}_i$ (equation 14) is $N - 1$, at most $N - 1$ subchannels can use irregular modulation irrespective of how large is the number of subchannels Q. Taking into account that usually $Q \gg N$, we can anticipate that the introduction of irregular modulation will not provide a significant MI gain in practical cases. Figure 8 depicts also the MI obtained by the simplified version of the proposed algorithm obtained after rounding α_{iq} coefficients to $\{0, 1\}$ to get rid of irregular modulation. As it can be seen, losses due to the rounding operation are negligible (the optimum and suboptimum curves overlap completely). However, irregular modulation is an essential point of the proposed method, since it guarantees the convexity of function $\bar{I}_o$ required by the mercury/waterfilling criterion and, therefore, it allows to derive an optimum bit loading and power allocation algorithm. In spite of this fact, this remark suggests that, once the optimum solution has been derived with the algorithm proposed in this paper, the implementation complexity can be reduced at the expense of minor losses by getting rid of irregular modulation and rounding the α_{iq} parameters to integers $\{0, 1\}$, so only one constellation is employed per subchannel.

Figure 7 also illustrates the scarce use of irregular modulation. It shows the optimum bit loading and power allocation per subchannel (P_q, equation 3) for an SNR of 8dB ($\gamma = 8dB, P_T = 1$) in the same scenario with 32 subchannels employed before. It depicts also the MWF power allocations in the case of using QPSK or 64QAM for all subchannels of the BICM scheme. For the available constellation set, this channel realization and this SNR, the optimum algorithm does not employ irregular modulation in any of the 32 subchannels (all α_{iq} parameters are either 0 or 1), so Fig. 7 depicts the only constellation that is employed in every subchannel. Furthermore, note that subchannel nr. 28 is allocated no power and constellation 256QAM is not employed. This procedure avoids the need to implement an irregular modulation scheme while the performance is kept nearly optimal in terms of mutual information.

It is worth mentioning that the proposed algorithm has not been found to have any special repercussion to the PAPR of the transmitted OFDM signal due to the inclusion of an irregular modulation scheme. Besides, considering a more realistic scenario in which the channel state information (CSI) is not perfect, it has been verified that the sensitivity of the adaptive design to this uncertainty is similar to that one of the waterfilling power allocation.

C. Algorithm Complexity

The complexity of the proposed algorithm is equivalent to the one of mercury/waterfilling power allocation. The algorithm implementation requires two steps: First, the the power allocated per subchannel is evaluated with mercury/waterfilling assuming that the equivalent constellation with mutual information given by equation (19) is employed. Second, the bit loading is obtained from a look-up table that stores the functions in equations (17) and (18).

The first stage, mercury/waterfilling, can be solved with superlinear convergence with algorithms such as the secant method [4, Appendix F], with the number of iterations depending on the desired tolerance and not the number of subchannels. In order to obtain the power allocation as defined in equation (21), the evaluation of the waterlevel (δ) as the root of the following function is required:

$$f(\delta) = \sum_{\substack{q=1 \\ \gamma_q > \delta}}^{Q} \frac{1}{\gamma_q} D_o^{-1}\left(\frac{\delta}{\gamma_q}\right) - Q \qquad (22)$$

This function has linear complexity with respect to the number of subchannels Q. It requires a maximum of $Q + 1$ summations, $2Q$ divisions and Q quantization operations (of the argument of D_o^{-1}, used to evaluate that function using a look-up table with no need of interpolation). Therefore, the maximum final cost is approximately $4nQ$ operations, where n is the number of iterations of the secant method.

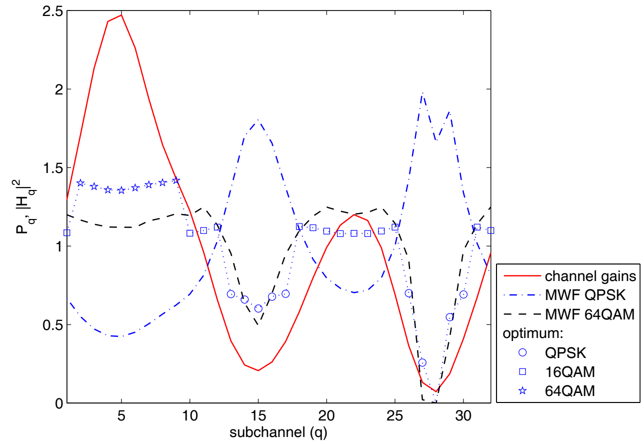


Fig. 7. Bit loading and power allocation for an example with 32 parallel subchannels. The allocated power according to the optimum algorithm is represented with a different marker depending on the constellation used per subchannel. The channel gains and the power allocated according to MWF with uniform bit loading are also represented.

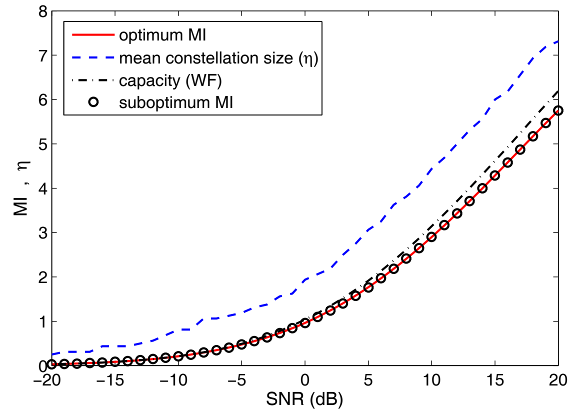


Fig. 8. MI versus SNR curve and mean constellation size for an example with 32 parallel subchannels.

The second stage, bit loading, has negligible complexity in the proposed algorithm. Note that the cost of bit loading algorithms in the literature can be significant, since iterative and greedy algorithms have non-linear complexity with respect to the number of subchannels. Simplified algorithms that take into account the quantization and make this dependence linear have been proposed, being Levin-Campello one of them [8].

VI. Conclusions

In this paper we propose an optimal bit loading and power allocation algorithm that maximizes the mutual information with a power constraint for parallel subchannels. This algorithm can be employed for performance optimization in BICM-OFDM systems with channel state information at the transmitter. As opposed to other bit loading algorithms in the literature, the proposed approach employs the constellation-constrained mutual information function rather than approximations of the throughput and does not result in a greedy algorithm.

The optimization can be decomposed in two steps. The first step consists of the design of an irregular modulation and power allocation for an AWGN channel and has a closed-form solution. The second step must be computed for each channel realization and it requires the application of the mercury/waterfilling power allocation over the set of parallel subchannels employing in all of them the equivalent constellation obtained in the first step. Assuming that the optimum mutual information resulting of the first step is tabulated, the complexity of our design is equivalent to the complexity of mercury/waterfilling, corresponding to the search of the optimum waterlevel. Hence, the additional computational load due to the incorporation of the bit loading is negligible.

The proposed approach relies on the application of irregular modulation and power allocation to obtain a convex problem formulation. However, in practical scenarios, the number of subchannels that employ symbols from more than one constellation is very small. Hence, nearly optimal performance can be obtained rounding the α_{iq} parameters (fraction of use) to either 0 or 1. This way, an algorithm for conventional bit loading and power allocation is obtained with nearly optimal performance and low complexity.

VII. Appendix A
Computation of the AWGN Channel Optimization Solution of Section IV

In this section we derive the solution to equations (10) provided in equations (17) and (18). First, let us note that, due to the definition of the s_i parameters introduced to cast the problem in the convex optimization framework, the points where $\alpha_i = 0$ are excluded from the possible solutions domain. However, this is not a restriction because for $\alpha_i \to 0$ (which means that the i-th constellation is not used) the corresponding term does not contribute to the mutual information:

$$\lim_{\alpha_i \to 0} \alpha_i I_i(\frac{s_i}{\alpha_i}\mu) = 0$$

since the mutual information is positive and upper bounded by m_i. Also, by definition, when the constellation is not used no power is allocated to it and vice versa, that is:

$$\alpha_i = 0 \Leftrightarrow p_i = 0$$

From the optimization problem (10), we define the Lagrangian function

$$J = -\sum_{i=1}^{N} \alpha_i I_i \left(\frac{s_i}{\alpha_i}\mu\right) - \sum_i \lambda_i \alpha_i + \lambda_0 \left(\sum_i \alpha_i - 1\right) - \sum_i \tau_i s_i + \tau_0 \left(\sum_i s_i - 1\right) \tag{23}$$

where the parameters $\{\lambda_i\}$ and $\{\tau_i\}$ are the Lagrange multipliers.

Since the optimization problem is convex, it has a unique solution which is given by the KKT conditions [15]. In our case these are:

$$\lambda_i \alpha_i = 0, \quad \lambda_i \geq 0 \quad \forall i \tag{24a}$$

$$\tau_i s_i = 0, \quad \tau_i \geq 0 \quad \forall i \tag{24b}$$

$$\lambda_0 (\textstyle\sum_i \alpha_i - 1) = 0, \quad \lambda_0 \geq 0 \tag{24c}$$

$$\frac{\partial J}{\partial s_i} = -\mu D_i \left(\frac{s_i}{\alpha_i}\mu\right) - \tau_i + \tau_0 = 0, \quad \forall i \tag{24d}$$

$$\frac{\partial J}{\partial \alpha_i} = -C_i \left(\frac{s_i}{\alpha_i}\mu\right) - \lambda_i + \lambda_0 = 0, \quad \forall i \tag{24e}$$

and also the initial constraints

$$\alpha_i \geq 0\,, \quad s_i \geq 0\,, \quad i = 1, ..., N \tag{25a}$$

$$\sum_{i=1}^{N} s_i = 1 \tag{25b}$$

$$\sum_{i=1}^{N} \alpha_i \leq 1 \tag{25c}$$

The optimization problem can be solved by finding the solution to this set of equalities and inequalities. To obtain a closed expression, we manipulate these equations as follows. From equation (24d),

$$\tau_i = \tau_0 - \mu D_i \left(\frac{s_i}{\alpha_i}\mu\right)$$

Then, introducing it on (24b),

$$\tau_0 = \mu D_i \left(\frac{s_i}{\alpha_i}\mu\right) \quad or \quad s_i = 0 \tag{26}$$

and also,

$$\tau_0 \geq \mu D_i \left(\frac{s_i}{\alpha_i}\mu\right)$$

Doing the same for the gradient with respect α_i, from equation (24e),

$$\lambda_i = \lambda_0 - C_i \left(\frac{s_i}{\alpha_i}\mu\right)$$

and then, on (24a),

$$\lambda_0 = C_i \left(\frac{s_i}{\alpha_i}\mu\right) \quad or \quad \alpha_i = 0 \tag{27}$$

and also

$$\lambda_0 \geq C_i \left(\frac{s_i}{\alpha_i}\mu\right) \tag{28}$$

Introducing (27) in (24c),

$$C_i \left(\frac{s_i}{\alpha_i}\mu\right) = 0 \quad or \quad \textstyle\sum_i \alpha_i = 1 \tag{29}$$

The first condition is only possible if all the quotients $\frac{s_i}{\alpha_i}$ are 0, that is, the power allocated to all constellations is zero and, coherently, no symbols are transmitted, which is not a valid solution. Therefore, the second condition is always fulfilled: the sum of α's is 1, what means that all the channel accesses are used to transmit a symbol.

Going back to equality (26) for $s_i \neq 0$, and introducing the equality in expression (27) for $\alpha_i \neq 0$, we obtain

$$\lambda_0 = L_i \left(\frac{\tau_0}{\mu}\right) \quad or \quad \alpha_i = 0, s_i = 0 \tag{30}$$

This establishes a set of equalities between the two Lagrange multipliers λ_0 and τ_0 that must be fulfilled for each constellation ($i = 1, ..., N$) that is employed (non-null parameters α_i and s_i). Moreover, doing the same for inequality (28)

$$\lambda_0 \geq L_i\left(\frac{\tau_0}{\mu}\right)$$

and, therefore,

$$\lambda_0 = max_i \quad L_i\left(\frac{\tau_0}{\mu}\right) \quad or \quad \alpha_i = 0, s_i = 0 \tag{31}$$

This equation leads to the definition of the regions

$$\mathcal{R}_i = \{x \quad : \quad L_i(x) \geq L_j(x) \quad \forall j \neq i\} \tag{32}$$

which where introduced in section IV (equation 14). The interior points of these regions correspond to pairs of values λ_0 and τ_0 where a unique constellation fulfills equation (31) for non-null allocated parameters. As seen in Fig. 2, the boundary points of these regions are the intersections between two of these curves, defined as ν_i in equation (15). This means that two constellations fulfill simultaneously equation (31) with non-null allocated parameters. Thus, we define two sets of possible hypothesis that determine the final solution:

1) $\frac{\tau_0}{\mu} \in \mathcal{R}_i, \quad \frac{\tau_0}{\mu} \neq \nu_{i-1}, \nu_i \quad i = 1, ..., N$: Only one constellation is used, when the corresponding Legendre transform is maximum. Then the solution is trivial: $\alpha_j = \delta_{ij}$ and $s_j = \delta_{ij}$[3] Then, from equation (26),
$$\frac{\tau_0}{\mu} = D_i(\mu) \tag{33}$$
and the condition for this hypothesis to be valid turns to: $D_i(\mu) \in \mathcal{R}_i, \quad D_i(\mu) \neq \nu_{i-1}, \nu_i$
2) $\frac{\tau_0}{\mu} = \nu_i \quad i = 1, ..., N-1$: Two constellations are combined, when their Legendre transforms intersect. Then the parameters s_i, s_{i+1}, α_i and α_{i+1} are not null and found from the constraints on their sum and equations (26). Rewriting them for this specific hypothesis, the following set of equations is obtained:

$$\alpha_{i+1} + \alpha_i = 1 \tag{34a}$$

$$s_{i+1} + s_i = 1 \tag{34b}$$

$$s_i = \frac{\alpha_i}{\mu} D_i^{-1}(\nu_i) \tag{34c}$$

$$s_{i+1} = \frac{\alpha_{i+1}}{\mu} D_{i+1}^{-1}(\nu_i) \tag{34d}$$

With straightforward manipulations, we can express the solution in terms of μ and ν_i. For example, from the second set of hypothesis, we can isolate α_i:

$$\alpha_i = \frac{D_{i+1}^{-1}(\nu_i) - \mu}{D_{i+1}^{-1}(\nu_i) - D_i^{-1}(\nu_i)} \tag{35}$$

and then obtain the remaining parameters from it.

Notice that we have reduced the dependence on the Lagrange multiplier τ_0 and the SNR μ to only the dependence on the latter. Therefore, the value of μ determines which one of the previous hypothesis is valid, i.e. with parameters fulfilling constraints (25). Indeed, we can introduce these constraints in the equations of the second hypothesis (equation 35) in order to find the thresholds that determine which is the valid one:

- First, $\alpha_i > 0$ and then,
$$\mu < D_{i+1}^{-1}(\nu_i) \tag{36}$$
- Second $\alpha_i < 1$ and then,
$$\mu > D_i^{-1}(\nu_i) \tag{37}$$

which are the thresholds defined in equation (16) at section IV.

The final expression of the solution in equations (17) and (18) is constructed in a piecewise manner, where each piece corresponds to one of the regions of SNR values where one of the $2N-1$ hypothesis is valid.

VIII. Appendix B
Low Complexity BICM Demapper Metrics

This appendix presents the bit metric computation for the low complexity BICM demapper considered in the simulations.

Let $(b_{q1}(n), ..., b_{qm}(n))$ denote the bits associated to the symbol $x_q(n)$ in equation (1). The bit log-likelihood ratios (LLR) that are delivered to the decoder are defined as

$$LLR\{b_{qj}(n)\} = log\frac{Pr\left(b_{qj}(n) = 1 | y_q(n)\right)}{Pr\left(b_{qj}(n) = 0 | y_q(n)\right)} \tag{38}$$

Omitting the temporal index n, this ratios are obtained as:

$$\begin{aligned} LLR\{b_{qj}\} = &\log \sum_{x_q : b_{qj}=1} \exp -\frac{1}{\sigma^2}|y_q - H_q\sqrt{p_q}x_q|^2 \\ &- \log \sum_{x_q : b_{qj}=0} \exp -\frac{1}{\sigma^2}|y_q - H_q\sqrt{p_q}x_q|^2 \end{aligned} \tag{39}$$

Applying the max-log approximation $log \sum_i \exp z_i \approx \max_i z_i$ as in [16], the LLR's are finally simplified as:

$$\begin{aligned} LLR\{b_{qj}\} \approx &\max_{x_q : b_{qj}=1} -\frac{1}{\sigma^2}|y_q - H_q\sqrt{p_q}x_q|^2 \\ &- \max_{x_q : b_{qj}=0} -\frac{1}{\sigma^2}|y_q - H_q\sqrt{p_q}x_q|^2 \end{aligned} \tag{40}$$

[3] δ_{ij} stands for the usual Kronecker delta, i.e. $\delta_{ij} = \begin{cases} 1 & if \quad j = i \\ 0 & otherwise \end{cases}$.

References

[1] E. Biglieri, G. Caire, and G. Taricco, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, pp. 927-946, May 1998.

[2] J. A. C. Bingham, "Multicarrier modulation for data transmission: an idea whose time has come," *IEEE Commun. Mag.*, vol. 28, no. 5, pp. 5-14, May 1990.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1990.

[4] A. M. Tulino, A. Lozano, and S. Verdu, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3033-3051, July 2006.

[5] G. D. Forney Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, pp. 2384-2415, Oct. 1998.

[6] P. S. Chow, J. M. Cioffi, and J. A. C. Bingham, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Trans. Commun.*, vol. 43, pp. 773-775, Apr. 1995.

[7] R. F. H. Fischer and J. B. Huber, "A new loading algorithm for discrete multitone transmission," in *Proc. Globecom*, London, U.K., 1996, pp. 927-946.

[8] J. Campello, “Practical bit loading for DMT," in *Proc. Int. Conf. Commun. (ICC’99)*, Vancouver, June 1999, vol. 2, pp. 801-805.

[9] B. S. Krongold, K. Ramchandran, and D. L. Jones, “Computationally efficient optimal power allocation algorithms for multicarrier communication systems," *IEEE Trans. Commun.*, vol. 48, pp. 23-27, Jan. 2000.

[10] N. Papandreou and T. Antonakopoulos, “A new computationally efficient discrete bit-loading algorithm for DMT applications," *IEEE Trans. Commun.*, vol. 53, pp. 785-789, May 2005.

[11] G. Bauch and F. Schreckenbach, “Bit-interleaved coded irregular modulation," *Europ. Trans. Telecommun.*, pp. 269-282, 2006.

[12] O. Lahuerta and M. Lamarca, “Joint design of code, constellation and power allocation in BICM LDPC coded systems," *4th Intl. Symp. Turbo Codes Related Topics*, 2006.

[13] D. Matas and M. Lamarca, “Bit loading and power allocation for BICM-ID with convolutional codes," *Signal Process. Advances Wireless Commun. (SPAWC’07)*, June 2007.

[14] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1996.

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

[16] F. Tosato and P. Bisaglia, “Simplified soft-output demapper for binary interleaved COFDM with application to Hiperlan/2," in *Proc. Int. Conf. Commun. (ICC’02)*, vol. 2, pp. 664-668.

[17] S. H. Han and J. H. Lee, “An overview of peak-to-average power ratio reduction techniques for multicarrier transmission," *IEEE Wireless Commun.*, vol. 12, pp. 56-65, Apr. 2005.

[18] “Broadband radio access networks (BRAN); HIPERLAN Type 2; physical (PHY) layer," ETSI, ETSI TS 101 475 v1.2.2, 2001.

[19] F. Rey, M. Lamarca, and G. Vazquez, “Robust power allocation algorithms for MIMO-OFDM systems with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 53, pp. 1070-1085, Mar. 2005.

David Matas (S’07) was born in Cambrils, Spain, in 1983. He received the M.S. degree in telecommunication engineering and the advanced studies diploma in signal theory and communications from the Technical University of Catalonia (UPC), Spain, in 2006 and 2009, respectively. Since 2006, he is a Research Assistant in the Department of Signal Theory and Communications of the UPC , where he holds a grant from the DIUE of the Catalan Government in support of his Ph.D. studies. His research interests are digital communications and information theory, with special attention to adaptive modulation and joint source-channel coding.

Meritxell Lamarca (M’97) was born in Barcelona, Spain, in 1967. She received the M.Sc. and Ph.D. degrees in telecommunication engineering from the Technical University of Catalonia (UPC), Barcelona, in 1992 and 1997, respectively. In 1992 she joined the Department of Signal Theory and Communications of the UPC and soon after she became a Lecturer. Since 1997 she has been an Associate Professor at UPC, and since 2009 she is on leave at the University of Delaware. Her general interests are signal processing and digital communications. She conducts research activities in coding and modulation for wireless communications, with particular emphasis in turbo detection and MIMO systems. She has been involved in many research projects in wireless and satellite communications under the research programs of the European Union and the European Space Agency.